

doi: 10.12052/gdutxb.240005

面向无人机数字孪生边缘网络的联邦学习资源 自适应优化机制

谢正昊^{1,2}, 赖健鑫^{1,2}, 庄晓翀^{1,3}, 蒋 丽^{1,2}

(1. 广东工业大学 自动化学院, 广东 广州 510006; 2. 广东工业大学 物联网信息技术广东省重点实验室, 广东 广州 510006; 3. 智能检测与制造物联教育部重点实验室, 广东 广州 510006)

摘要: 为了解决无人机数字孪生边缘网络联邦学习性能优化问题, 本文提出一种基于深度强化学习的无人机数字孪生边缘网络资源调度策略。考虑动态时变的无人机数字孪生边缘网络环境, 构建包含地面基站(Base Station, BS)、地面智能终端、空中无人机以及无线传输信道的孪生网络模型, 建立联合无人机飞行距离、飞行角度以及无线网络频谱资源分配的自适应资源优化模型, 实现最小化联邦学习时延的目标。在无人机数字孪生边缘网络环境下, 提出多智能体深度确定性策略梯度算法(Multi-Agent Deep Deterministic Policy Gradient, MA-DDPG), 求解自适应资源优化模型。算法训练过程采用中心化训练、去中心化执行的方式, 每个无人机智能体在评估动作价值时会考虑其他智能体的状态和动作, 而在执行时只根据自身的局部观察来决定动作。上述训练过程将在数字孪生环境中执行, 算法收敛后再应用于真实世界, 最大限度地减少物理实体的资源开销。仿真结果表明, 所提算法可显著降低联邦学习服务时延, 同时保证联邦学习训练损失和准确率的优越性。

关键词: 无人机网络; 数字孪生; 联邦学习; 多智能体深度确定性策略梯度

中图分类号: TN929.5

文献标志码: A

文章编号: 1007-7162(2024)04-0061-09

Adaptive Resource Optimization for Federated Learning in UAV Digital Twin Edge Networks

Xie Zheng-hao^{1,2}, Lai Jian-xin^{1,2}, Zhuang Xiao-chong^{1,3}, Jiang Li^{1,2}

(1. School of Automation, Guangdong University of Technology, Guangzhou 510006, China; 2. Guangdong Key Laboratory of IoT Information Technology, Guangdong University of Technology, Guangzhou 510006, China; 3. Key Laboratory of Intelligent Detection and the Internet of Things in Manufacturing, Ministry of Education, Guangzhou 510006, China)

Abstract: To address the performance optimization issues in federated learning for unmanned aerial vehicle (UAV) digital twin edge networks, a resource scheduling strategy is proposed based on deep reinforcement learning for UAV digital twin edge networks. Considering dynamic and time varying features of UAV digital twin edge networks environment, a twin network model is built consisting of base station (BS), intelligent terminals, UAV and wireless transmission channel. Then an adaptive resource optimization model is formulated which jointly optimized UAV flying distance, flying angle and spectrum resource allocation, in order to minimize time delay of federated learning. Moreover, an UAV digital twin edge networks empowered multi-agent deep deterministic policy gradient (MA-DDPG) algorithm is designed to solve the adaptive resource optimization model. The algorithm training process adopts centralized training and decentralized execution. Each UAV agent considers the states and actions of other agents when evaluating the value of actions, but decides actions based only on its own local observations during execution. The above training process is conducted in digital twin environment, and after the algorithm converges, and it is applied to the real world, minimizing the resource overhead of physical entities to the greatest extent. Numerical results illustrate that the proposed algorithm can significantly decrease the service

收稿日期: 2024-01-29

基金项目: 国家重点研发计划项目(2020YFB1807801); 国家自然科学基金资助面上项目(62371142, 62273107)

作者简介: 谢正昊(1998-), 男, 硕士研究生, 主要研究方向为无人机网络、联邦学习和数字孪生等, E-mail: redteaice@foxmail.com

通信作者: 蒋丽(1986-), 女, 副教授, 主要研究方向为6G网络和网络内生安全等, E-mail: jiangli@gdut.edu.cn

latency of federated learning, while guaranteeing the superiority of training loss and accuracy of federated learning.

Key words: unmanned aerial vehicle networks; digital twin; federated learning; multi-agent deep deterministic policy gradient

由于无人机具有移动性可控、灵活部署以及视距通信等特点,无人机可为物联网边缘应用提供服务^[1-3],例如无人机可以作为空中基站(Base Station, BS)或者中继节点,实现地面智能终端无处不在的连接。因此,无人机网络通过协调多个无人机通信,可以实现自主的空中网络,已受到学术界和工业界广泛关注。然而,随着无人机网络中地面智能终端设备激增、多场景多业务动态叠加、网络规模持续扩大以及无人机快速移动等特征,无人机网络优化和管理越来越复杂。数字孪生采用建模工具构建无人机网络中地面BS、地面智能终端、空中无人机以及无线传输信道等物理实体的高保真数字模型(即数字孪生体)。在未来6G无人机网络时代,大量无人机网络数字孪生体将被部署在网络边缘服务器上,通过虚实交互,保持无人机网络物理实体和数字孪生体状态同步更新和协同演化,最终形成无人机数字孪生边缘网络,并采用分布式智能算法,为无人机网络优化和管理提供智能化决策,更加高效地应对无人机网络复杂挑战^[4-6]。

然而,无人机数字孪生边缘网络在协同边缘服务器上的数字孪生体进行智能化决策时,边缘服务器需要获取海量异构数据,这些数据涉及智能终端敏感信息,例如生物特征识别信息、车辆自动驾驶控制信息等。多个互不信任的数字孪生体协同计算,容易泄露智能终端敏感信息,导致无人机数字孪生边缘网络安全问题。联邦学习作为一种新兴的分布式机器学习架构,智能终端在本地训练机器学习模型,并将训练好的本地模型参数上传到中心云服务器聚合,而不用上传原始数据,从而可有效保护智能终端数据安全和隐私^[7]。目前,联邦学习已经应用于计算机视觉、自然语言处理、汽车自动驾驶、智能家居等领域^[8]。当前联邦学习应用于无人机数字孪生边缘网络,仍面临严峻挑战。无人机有限的计算资源、存储能力、能量资源、高速移动性,以及动态时变的无线传输环境,将导致联邦学习模型参数传输出错的概率增大,进而使联邦学习性能下降^[9]。Song等^[10]提出基于非正交多接入技术的无人机网络联邦学习架构,通过联合优化上行非正交多接入传输时延、下行广播时延以及无人机的计算时延,实现最小化联邦学习时延的目标。Cheng等^[11]研究无人机网络多任务

联邦学习激励机制,提出使用拍卖博弈算法激励无人机和智能终端参与联邦学习模型训练。Sun等^[12]提出采用数字孪生辅助无人机网络联邦学习激励机制和斯塔克伯格算法,激励地面智能终端贡献资源参与联邦学习模型训练,提高联邦学习准确率。

考虑无人机数字孪生边缘网络动态时变的无线传输环境,当前大部分研究将深度强化学习(Deep Reinforcement Learning, DRL)算法和无人机网络的孪生网络模型结合,DRL算法可以从孪生网络模型中获取反映真实无人机网络状态的信息,并利用这些状态信息进行模型训练。同时,孪生网络模型实时获取真实无人机网络状态变化信息,且连续地更新模型参数,以对真实无人机网络的性能优化做出最优决策。Li等^[13]提出采用数字孪生辅助无人机网络边缘计算任务卸载,将数字孪生和双层深度Q网络(Double Deep Q-Network, DDQN)算法结合,求解智能终端最优关联和无人机最优飞行轨迹,实现最小化无人机网络边缘计算能耗的目标。Chen等^[14]提出无人机网络智能终端任务卸载调度策略,将数字孪生和DRL算法结合,在构建的孪生网络模型中训练智能终端任务卸载调度决策模型,实现最大化无人机网络地面智能终端平均效用的目标。这些工作聚焦于无人机数字孪生边缘网络中计算任务卸载策略设计,然而在无人机数字孪生边缘网络的联邦学习中,如何应对复杂变化的网络环境,并协同多个无人机提升联邦学习性能仍存在极大挑战。与现有工作不同,本文在无人机数字孪生边缘网络中,设计无人机多智能体协作调度策略,以降低联邦学习时延,同时保证学习准确率和损失的优越性。因此本文主要贡献如下:

(1) 提出基于无人机数字孪生边缘网络的联邦学习资源调度策略。首先,采用数字孪生技术在地面BS构建无人机网络的孪生网络模型,对无人机网络运行状态进行模拟和仿真。考虑多个无人机作为联邦学习模型参数的聚合器,对地面智能终端训练的本地模型参数进行聚合,在保护地面智能终端原始数据安全的前提下提供分布式机器学习服务。

(2) 进一步将无人机数字孪生边缘网络联邦学习资源分配问题建模为马尔科夫决策过程,建立联合无人机飞行距离、飞行角度以及无线网络频谱资

源分配的自适应资源优化模型,实现最小化联邦学习时延的目标。提出基于孪生网络模型的多智能体深度确定性策略梯度(Multi-Agent Deep Deterministic Policy Gradient, MA-DDPG)的资源分配优化算法,求解自适应资源优化模型。

(3) 仿真结果表明,本文算法具有良好的收敛性。与其他算法对比,本文算法可显著降低联邦学习时延性能,同时保证了联邦学习训练损失和准确率的优越性。

1 系统模型

本文构建的无人机数字孪生边缘网络联邦学习架构如图1所示,由真实无人机网络模型和孪生网络模型组成。真实无人机网络模型中, N 个智能终端部署在区域 $\mathcal{A}$ 中。同时,考虑集合 $\mathcal{M} = \{1, 2, \dots, M\}$ 的无人机配备有边缘服务器,地面智能终端向无人机发送辅助计算需求。在无人机对地面智能终端计算服务需求充分认知的情况下,为地面用户提供高质量计算服务,如车辆自动驾驶预警、导航等。每个无人机的覆盖区域用 $\mathcal{A}_m$ 表示,且对于任意 $m, m' \in \mathcal{M}$, $m \neq m'$, 存在 $\cup_{m \in \mathcal{M}} \mathcal{A}_m = \mathcal{A}$ 且 $\mathcal{A}_m \cap \mathcal{A}_{m'} = \emptyset$ 。在无人机 m 的区域 $\mathcal{A}_m$ 中,分布有 N_m 个智能终端,用集合 $\mathcal{N}_m = \{1, 2, \dots, N_m\}$ 表示。因此,区域 $\mathcal{A}$ 中分布的所有智能终端集合为 $\mathcal{N} = \{1, 2, \dots, N\}$,且 $\mathcal{N} = \cup_{m \in \mathcal{M}} \mathcal{N}_m$ 。为了提高计算服务质量,每个无人机只为覆盖区域内的智能终

端提供计算服务。由于无人机的电池能量有限,设每个无人机最大服务时长为 T ,设其在满电状态且亏电前不补充电量前提下的最大飞行距离为 $d_m^{\max}$ 。设在无人机提供计算服务过程中,始终保持在高度为 H 的水平面上。

在时刻 $t \in \{0, 1, 2, \dots, T\}$,考虑智能终端 n 的数据表示为 $\mathcal{D}_n = \{(x_{n1}, y_{n1}), \dots, (x_{nD_n}, y_{nD_n})\}$,且无人机的数据表示为 $\mathcal{D}_m = \{(x_{m1}, y_{m1}), \dots, (x_{mD_m}, y_{mD_m})\}$ 。其中, D_n 表示智能终端 n 的数据大小, D_m 表示无人机 m 的数据大小。地面BS连续地采集服务区域 $\mathcal{A}$ 内智能终端和无人机的数据,以构建包含智能终端、无人机、网络拓扑及环境的孪生网络模型。智能终端 n 和无人机 m 在地面BS的孪生模型分别表示为 $\text{DT}_n = \{\mathcal{M}_n, \mathcal{D}_n, S_{n,t}, S_{n,t+1}\}$ 和 $\text{DT}_m = \{\mathcal{M}_m, \mathcal{D}_m, S_{m,t}, S_{m,t+1}\}$ 。其中, $\mathcal{M}_n$ 和 $\mathcal{M}_m$ 分别表示智能终端 n 和无人机 m 的行为模型, $S_{n,t}$ 和 $S_{m,t}$ 分别表示智能终端 n 和无人机 m 的实时状态, $S_{n,t+1}$ 和 $S_{m,t+1}$ 分别表示智能终端 n 和无人机 m 的更新状态。地面BS通过生成的孪生网络模型对无人机网络运行状态进行模拟和仿真,并使用合适的智能算法设计无人机边缘计算方案,以满足地面用户的计算服务性能需求。

在本文提出的系统模型中,无人机充当联邦学习的聚合器,聚合维护具体计算任务的全局模型参数。地面智能终端从服务区域的无人机获得全局模型参数,在全局模型的基础上使用本地数据训练本地模型;然后智能终端通过上行信道将训练后的本地模型参数上传到服务区域的无人机,由无人机对智能终端本地模型参数进行聚合以更新全局模型参数。

1.1 联邦学习模型

用 w_m 表示第 m 个无人机的全局模型参数,用 w_n 表示第 n 个智能终端的本地模型参数,用 $f(w_m, x_{n,i}, y_{n,i})$ 表示关于数据样本 $(x_{n,i}, y_{n,i})$ 和参数 w_m 的损失函数,则智能终端 n 的本地模型参数训练损失可以表示为

$$f_n(w_m) = \frac{1}{D_n} \sum_{i \in \mathcal{D}_n} f(w_m, x_{n,i}, y_{n,i}) \quad (1)$$

第 m 个无人机的全局模型参数训练损失表示为

$$f(w_m) = \frac{1}{D_m} \sum_{n \in \mathcal{N}_m} \sum_{i \in \mathcal{D}_n} f(w_m, x_{n,i}, y_{n,i}) \quad (2)$$

第 m 个无人机执行联邦学习,寻找最优模型参数,以最小化全局模型训练损失。

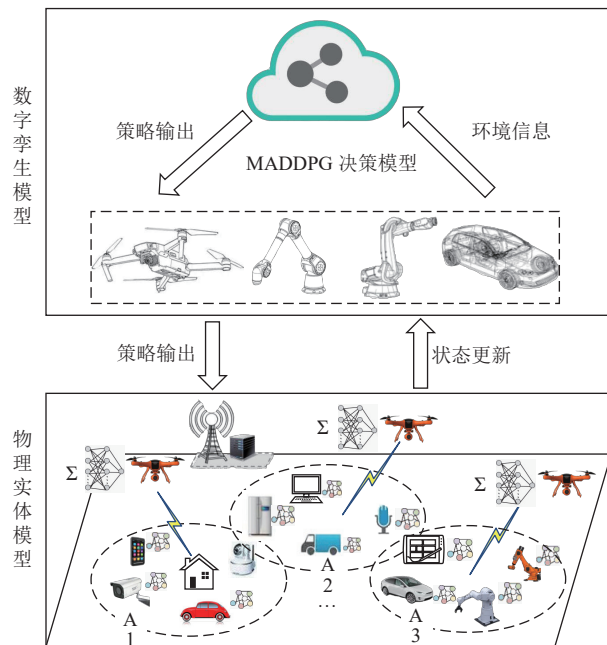


图1 无人机数字孪生边缘网络联邦学习架构

Fig.1 Federated learning architecture for drone digital twin edge networks

$$w_m^* = \operatorname{argmin}_{w_m} f(w_m) \quad (3)$$

1.2 通信模型

每一轮联邦学习训练过程中,考虑智能终端采用正交频分多址接入技术,通过上行信道将训练的本地模型参数上传到服务区域的无人机。区域 $\mathcal{A}$ 中有 K 个上行子信道,表示为 $\kappa = \{1, 2, \dots, K\}$ 。 N 个智能终端复用 K 个上行子信道。此时,第 m 个无人机将收到复用相同子信道的其他区域内智能终端干扰。因此,在时刻 t ,第 n 个智能终端与第 m 个无人机在子信道 k 上的信干噪比可以表示为

$$\text{SINR}_{n,m,k}^m = \frac{P_{n,k}^{\text{up}} g_{n,m}}{\sum_{n' \in \mathcal{N}, n' \neq n} P_{n',k}^{\text{up}} g_{n',m} + \sigma^2} \quad (4)$$

式中: $P_{n,k}^{\text{up}}$ 为智能终端 n 在子信道 k 上的发送功率, σ^2 是高斯噪声功率, $\sum_{n' \in \mathcal{N}, n' \neq n} P_{n',k}^{\text{up}} g_{n',m}$ 表示来自其他区域智能终端在子信道 k 的干扰; $g_{n,m}(t)$ 为智能终端 n 和无人机 m 之间的信道增益

$$g_{n,m}(t) = \frac{\beta}{d_{n,m}^2(t) + H^2} \quad (5)$$

式中: $d_{n,m}(t) = \sqrt{(X_m(t) - x_n)^2 + (Y_m(t) - y_n)^2}$ 为智能终端 n 和无人机 m 之间的水平投影距离。无人机 m 在时刻 t 的物理位置为 $\psi_m(t) = (X_m(t), Y_m(t), H)$,智能终端 n 的物理位置为 $\omega_n(t) = (x_n, y_n)$,考虑 M 个无人机的飞行高度均为 H 。因此,在时刻 t ,智能终端 n 在子信道 k 上的数据传输速率表示为

$$R_{n,m}(t) = B \sum_{k=1}^K \phi_{n,m,k}(t) \log_2(1 + \text{SINR}_{n,m,k}^m) \quad (6)$$

式中: B 为子信道的带宽, $\phi_{n,m,k}(t)$ 为子信道 k 分配决策变量,且 $\phi_{n,m,k} \in (0, 1)$ 。当 $\phi_{n,m,k} = 1$ 时,表示智能终端 n 选择子信道 k 上传本地模型参数,否则 $\phi_{n,m,k} = 0$ 。

1.3 计算模型

在联邦学习训练过程中,地面智能终端 n 根据本地数据样本训练本地模型参数,贡献的计算资源表示为 f_n^{CPU} ,即CPU周期频率。训练一个数据样本需要的CPU周期频率表示为 c_n ,则在本地训练过程中,智能终端 n 本地训练的计算时延表示为

$$T_{n,\text{loc}} = \frac{D_n c_n}{f_n^{\text{CPU}}}, \quad \forall n \in \mathcal{N} \quad (7)$$

本地模型参数训练完成后,智能终端将本地模型参数无线传输到服务区域的无人机,本地模型参数上传时延表示为

$$T_{n,m,\text{up}}(t) = \frac{|w_m|}{R_{n,m}(t)}, \quad \forall n \in \mathcal{N} \quad (8)$$

式中: $|w_m|$ 为本地模型参数大小, $R_{n,m}(t)$ 为智能终端 n 的数据上传速率。

每轮联邦学习训练总时延由本地模型训练时延、本地模型上传时延、全局模型聚合时延和全局模型广播时延组成。考虑无人机配备的边缘服务器有充足的计算资源,且无人机使用专门的信道进行全局模型参数广播。与本地模型训练和上传时延相比,无人机全局模型聚合时延和全局模型广播时延较小。因此,第 n 个智能终端参与一轮联邦学习训练的时延表示为

$$T_n(t) = T_{n,\text{loc}} + T_{n,m,\text{up}}(t), \quad \forall n \in \mathcal{N} \quad (9)$$

每轮联邦学习训练总时延由所有终端参与本轮联邦学习的最大时延决定,表示为

$$T_m(t) = \max T_n(t), \quad \forall n \in \mathcal{N}, m \in \mathcal{M} \quad (10)$$

为此,通过优化无人机的飞行距离、飞行角度、子信道分配,设计式(11)最小化无人机数字孪生边缘网络联邦学习时延的优化问题

$$\begin{aligned} & \min_{\alpha, d, \phi} \frac{1}{T} \sum_{t=1}^T T_m(t), \quad \forall m \in \mathcal{M} \\ \text{s.t.} \quad & \begin{cases} \text{C1: } 0 \leq X_m(t) \leq X_{\max}, 0 \leq Y_m(t) \leq Y_{\max}, \\ \quad \forall m \in \mathcal{M}, \\ \text{C2: } 0 \leq d_m(t) \leq d_m^{\max}, 0 \leq \varphi_m(t) \leq 2\pi, \forall m \in \mathcal{M}, \\ \text{C3: } d_{m,m'}(t) \geq d_{\min}, \quad \forall m \in \mathcal{M}, \\ \text{C4: } d_{n,m}(t) \leq R_{\max}, \quad \forall n \in \mathcal{N}, \forall m \in \mathcal{M}, \\ \text{C5: } \sum_{n \in \mathcal{N}_m} \sum_{k \in K} \phi_{n,m,k}(t) \leq K, \phi_{n,m,k}(t) \in \{0, 1\}, \\ \quad \forall n \in \mathcal{N}, \forall m \in \mathcal{M} \end{cases} \end{aligned} \quad (11)$$

式中:C1表示无人机不能飞出服务区域 $\mathcal{A}$;C2表示无人机在 t 时刻执行决策的飞行距离 $d_m(t)$ 不能超过最大飞行距离 $d_m^{\max}$,且飞行角度 $\varphi_m(t)$ 不能超过最大飞行角度 2π ;C3表示无人机 m 和 m' 之间的距离 $d_{m,m'}(t) = \sqrt{(X_m(t) - X_{m'}(t))^2 + (Y_m(t) - Y_{m'}(t))^2}$ 需大于最小碰撞距离 $d_{\min}$;C4为地面智能终端到无人机的水平投影距离小于无人机的覆盖区域半径 $R_{\max}$;C5为子信道最大分配数量约束。

2 无人机数字孪生边缘网络联邦学习资源自适应优化算法设计

由于无人机数字孪生边缘网络无线传输环境的随机性,以及网络边缘计算、通信等资源的动态时变性,用传统的优化方法较难求解优化问题(11)。MA-DDPG中每个智能体结合全部智能体的策略进行参

数更新,从而更适合用于处理多个智能体的复杂协作问题,并具有较高鲁棒性。此外,MA-DDPG采用中心化训练、去中心化执行框架,允许智能体使用其他智能体的策略信息来增强自身的训练效果,而在实际应用或测试时不需要全局信息也能做出有效决策。Ryan等^[15]通过与已有算法在协作和竞争场景下的比较,证明了MA-DDPG算法性能优越性。

本文设计基于孪生网络模型的MA-DDPG优化算法,以求解无人机数字孪生边缘网络联邦学习时延优化问题(11)。如图2所示,在设计的优化算法中,MA-DDPG算法的运行环境由本地BS维护的孪生网络模型组成,包括无人机、地面智能终端以及无线传输信道的孪生模型。本文所提模型的具体流程如下:

步骤1:无人机网络中的地面智能终端和无人机将状态更新信息上传给本地BS维护的孪生网络模型,保持虚实同步更新。

步骤2:本地BS从孪生网络模型中观测无人机和地面智能终端的状态以及系统返回的即时报酬。

步骤3:将步骤2新观测到的状态输入MA-DDPG决策模型求解优化问题(11),获得输出决策。该决策会与状态和报酬等信息一起保存到MA-DDPG的经验缓存池中。当经验缓存池内的经验数据达到一定数量后,MA-DDPG开始迭代训练更新自身的网络参数。

步骤4:将MA-DDPG输出的决策应用到孪生网络模型中进行测试,孪生网络模型更新自身状态并计算即时报酬。

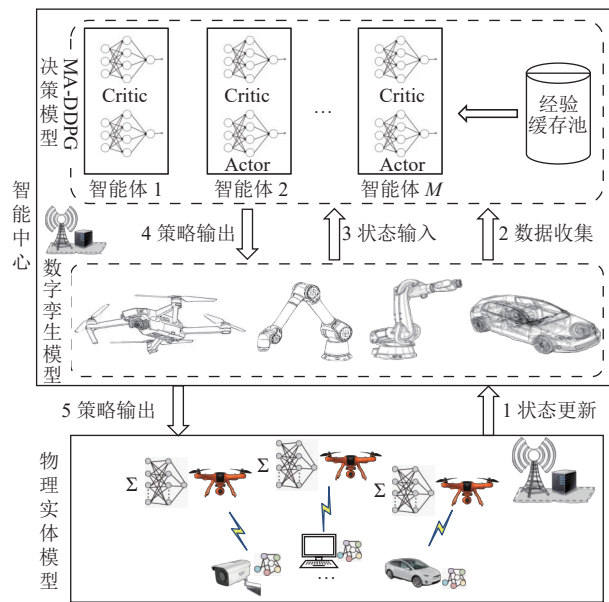


图2 基于孪生网络模型的MA-DDPG算法

Fig.2 MA-DDPG algorithm based on twin network model

步骤5:将经过测试的优化决策反馈给无人机和地面智能终端的物理实体执行。

相比传统MA-DDPG算法,本文设计的基于孪生网络模型的MA-DDPG算法,本地BS从孪生网络模型中获取状态信息,可减少MA-DDPG算法训练过程中无人机和地面智能终端交互的资源开销。

2.1 基于孪生网络模型的MA-DDPG优化算法

本文提出的基于孪生网络模型的MA-DDPG优化算法中,将决策问题构造造成马尔可夫决策过程,其中,状态空间、动作空间和报酬函数分别定义如下。

(1) 状态空间。在决策时刻 $t, t \in \{0, 1, \dots, T\}$,本地BS维护的孪生网络模型状态空间包括:无人机 m 的位置 $\psi_m(t)$,无人机 m 和无人机 m' 之间的距离 $d_{m,m'}(t)$,地面智能终端 n 的位置 $\omega_n(t)$,地面智能终端联邦学习的本地模型参数 $w_n(t)$,以及地面智能终端 n 和无人机 m 之间的信道增益 $g_{n,m}(t)$,其中, $n \in \mathcal{N}, m \in \mathcal{M}$ 。因此,状态空间可以表示为

$$S_{m,t} = \{\psi_m(t), d_{m,m'}(t), \omega_n(t), w_n(t), g_{n,m}(t)\} \quad (12)$$

(2) 动作空间。为了最小化无人机数字孪生边缘网络联邦学习时延,需要调整决策变量以适应网络环境的动态时变性。本地BS需要调整的决策变量包括:无人机在第 t 时刻的飞行距离 $d_m(t)$,无人机飞行角度 $\varphi_m(t)$,地面智能终端 n 的子信道分配决策 $\phi_{n,m,k}(t)$ 。其中, $d_m(t)$ 和 $\varphi_m(t)$ 可以调整无人机的位置以降低式(4)中无人机与地面智能终端交互时的信噪比 $\text{SINR}_{n,m,k}^m$, $\phi_{n,m,k}(t)$ 通过子信道的分配情况来调整式(6)中地面智能终端上传模型的速率 $R_{n,m}(t)$ 。上述3个决策变量相互配合可降低式(8)中所有地面智能终端本地模型参数的上传时间 $T_{n,m,\text{up}}(t)$,降低式(10)中的无人机 m 服务的 $T_m(t)$ 。所有无人机执行上述决策配合训练,每个无人机均实现其服务的联邦学习训练总时延 $T_m(t)$ 最小化,进而解决本文的优化问题(11)。因此,在决策时刻 t ,优化无人机数字孪生边缘网络联邦学习时延的动作空间表示为

$$A_{m,t} = \{d_m(t), \varphi_m(t), \phi_{n,m,k}(t)\} \quad (13)$$

(3) 报酬函数。考虑到强化学习以最大化长期平均报酬为目标,而本文需要最小化联邦学习的训练时延,因此设计如下即时报酬函数 $r_{m,t}$

$$r_{m,t} = \frac{l}{T_m(t)} \quad (14)$$

式中: l 为大于0的常数。MA-DDPG智能体的优化目标是最大化长期平均报酬

$$\mathcal{R} = \max \mathbb{E} \left[\sum_{t=0}^{T-1} \sum_{m=1}^M \gamma \cdot r_{m,t}(S_{m,t}, A_{m,t}) \right] \quad (15)$$

式中: γ 是折扣因子。

2.2 MA-DDPG算法原理

DDPG采取确定性策略,以第 m 个无人机智能体为例,其Actor作为策略网络,以环境状态 $S_{m,t}$ 作为输入,输出一个确定性动作 $A_{m,t}$ 。为保证神经网络的泛化能力且防止过拟合,DDPG会在策略网络Actor输出动作时添加随机噪声 o ,增加智能体尝试探索潜在最优策略的概率,即 $A_{m,t} = \mu_{\theta_m^{\text{Actor}}}(S_{m,t}) + o$,其中 $\mu_{\theta_m^{\text{Actor}}}(S_{m,t})$ 是Actor网络的输入与输出的关系表示, θ_m^{Actor} 是Actor网络的参数。第 t 时刻 M 个无人机智能体所处的环境状态集合表示为 $S_t = \{S_{1,t}, S_{2,t}, \dots, S_{M,t}\}$,动作集合表示为 $A_t = \{A_{1,t}, A_{2,t}, \dots, A_{M,t}\}$,即时报酬集合表示为 $r_t = \{r_{1,t}, r_{2,t}, \dots, r_{M,t}\}$ 。Target Actor是一个与Actor具有相同结构但不同参数的缓慢更新网络,将下一时刻的环境状态 $S_{m,t+1}$ 作为输入,输出与 $A_{m,t}$ 具有相同形状的动作 $A'_{m,t+1}$,但该输出仅用于更新价值网络Critic。MA-DDPG中每个智能体的价值网络Critic的作用是拟合状态-动作值函数 $Q_m(S_t, A_{m,t})$,以 M 个无人机智能体的环境状态集合 S_t 和对应第 m 个智能体的策略动作 $A_{m,t}$ 为输入,输出该动作于当前状态的价值 q -value。同样地,Target Critic与Critic具有相同的网络结构,以 S_{t+1} 、 $A'_{m,t+1}$ 为输入,生成对Target Actor网络输出 $A'_{m,t+1}$ 的评估价值 $Q_m^{\text{Target}}(S_{t+1}, A'_{m,t+1})$ 。DDPG借鉴了深度Q网络(Deep Q Network, DQN)的经验回放技巧,同样会把 (S_t, A_t, r_t, S_{t+1}) 存储到经验回放池。DDPG每次训练并更新网络参数时,会从经验回放池随机采样一个minibatch大小的经验作为本次训练的数据样本。

MA-DDPG的基本思想是中心化训练、去中心化执行。每个无人机智能体可利用全局信息,包括其他无人机智能体所处的环境状态及对应的策略动作等信息。对Critic进行训练,但在其进行决策时只需局部信息(该智能体自身所处的环境状态),就可得到确定性策略。第 m 个无人机智能体会独立地学习集中式的状态-动作值函数 $Q_m(S_t, A_{m,t})$,因此可通过为每个智能体设置不同报酬函数来实现不同智能体间的协作。

价值网络Critic采用基于TD-error的梯度下降更新方法,该网络需要结合环境的真实收益 $r_{m,t}$ 和 $Q_m^{\text{Target}}(S_{t+1}, A'_{m,t+1})$ 来学习如何对策略网络Actor的动作价值做出准确的评价。本文所提出的优化方案中,

所有无人机智能体之间是相互协作的关系,目标是使所有无人机智能体的平均报酬最大化。因此,第 m 个无人机智能体的Critic网络损失函数表示为

$$L_m^{\text{Critic}}(\theta_m^{\text{Critic}}) = \text{MSE}(Q_m(S_t, A_{m,t}), \frac{1}{M} \sum_{m=1}^M r_{m,t} + \gamma Q_m^{\text{Target}}(S_{t+1}, A'_{m,t+1})) \quad (16)$$

策略网络Actor的目标是找到当前环境状态下的最优策略动作,优化策略网络Actor就需要最大化价值网络输出 $Q_m(S_t, A_{m,t})$ 。因此,Actor网络采用梯度上升的更新方法,第 m 个无人机智能体的损失函数表示为

$$L_m^{\text{Actor}} = -Q_m(S_t, A_{m,t}) \quad (17)$$

第 m 个无人机智能体的累计期望报酬

$$J(\theta_m^{\text{Actor}}) = E_{S_{m,t}, A_{m,t} \sim \mu_{\theta_m^{\text{Actor}}}} \left[\sum_{t=0}^{\infty} \gamma \cdot \frac{1}{M} \sum_{m=1}^M r_{m,t} \right] \quad (18)$$

对应第 m 个无人机智能体的确定性策略 $\mu_{\theta_m^{\text{Actor}}}$,其梯度为

$$\nabla_{\theta_m^{\text{Actor}}} J(\mu_{\theta_m^{\text{Actor}}}) = E_{S_{m,t}, A_{m,t} \sim D} [\nabla_{\theta_m^{\text{Actor}}} \mu_{\theta_m^{\text{Actor}}}(S_{m,t}) \nabla_{A_{m,t}} Q_m(S_t, A_{m,t})] \quad (19)$$

式中: D 为元素 (S_t, A_t, r_t, S_{t+1}) 组成的经验回放池。

Target Actor的网络参数 θ_m^{Actor} 以及Target Critic的网络参数 θ_m^{Critic} 则分别按下式进行更新

$$\theta_m^{\text{Actor}} = \tau \theta_m^{\text{Actor}} + (1 - \tau) \theta_m^{\text{Actor}} \quad (20)$$

$$\theta_m^{\text{Critic}} = \tau \theta_m^{\text{Critic}} + (1 - \tau) \theta_m^{\text{Critic}} \quad (21)$$

其中 τ 用于调节Target网络更新时的参数占比。算法1给出了本文提出的MA-DDPG算法原理。

算法1 MA-DDPG算法原理

输入 状态 $S_1, S_2, \dots, S_T$

输出 对应时刻 t 所有无人机代理的决策 $A_1, A_2, \dots, A_T$

(1) 当episode=1,2,...,循环1:

(2) 对于 $t=1,2,\dots,T$,循环2:

(3) M 个无人机分别从孪生网络模型中获取当前时刻 t 的状态 $S_{m,t}, m \in \mathcal{M}$,并输入各自的Actor网络得到决策 $A_{m,t} = \mu_{\theta_m^{\text{Actor}}}(S_{m,t}) + o, m \in \mathcal{M}$

(4) M 个无人机分别在各自的孪生网络模型中执行决策,并得到各自的实时报酬及下一时刻的状态

(5) 存储 (S_t, A_t, r_t, S_{t+1}) 到经验回放池

(6) 对于代理=1,2,..., $\mathcal{M}$,循环3:

(7) 从经验回放池中随机采样一个minibatch大小的经验

- (8) 根据式(17)、(19)分别更新参数 θ_m^{Actor} 、 θ_m^{Critic}
- (9) 根据式(20)、(21)分别更新参数 θ_m^{Actor} 、 θ_m^{Critic}
- (10) 结束循环3
- (11) 结束循环2
- (12) 结束循环1

3 仿真结果

本节通过实验验证提出的基于孪生模型的MA-DDPG优化算法的性能。利用Pytorch神经网络框架搭建MA-DDPG优化算法。考虑在长和宽均为500 m的正方形区域 $\mathcal{A}$ 内部署4架无人机,无人机的飞行高度保持100 m,覆盖区域为半径 $R_{\max} = 10$ m的圆形,每次的最大飞行距离 $d_m^{\max} = 10$ m,飞行速度为10 m/s。区域 $\mathcal{A}$ 内上行子信道数量 $K=50$,每个子信道的上传带宽为10 MHz。区域 $\mathcal{A}_m$ 内的地面智能终端数量 $\mathcal{N}_m \in [20, 60]$,第 n 个终端设备可贡献的计算资源 $f_n^{\text{CPU}} \in [1.5 - 3.0]$ GHz,用于计算长期平均报酬的折扣因子 $\gamma = 0.97$ 。在MA-DDPG算法实现过程中,设置Actor和Critic网络分别由1个输入层,2个隐藏层和1个输出层构成。此外,本文还利用真实的数据集CIFAR10^[16]验证无人机数字孪生边缘网络联邦学习的性能,CIFAR10数据集由10种类别的60 000张图像组成,包括50 000张训练图像和10 000张测试图像。在进行仿真验证时,CIFAR10数据集被随机分配给地面智能终端。本文采用卷积神经网络作为联邦学习的机器学习模型。

图3对比了不同算法下无人机智能体的平均报酬。从图中可以看出,随着训练次数增加,不同算法下无人机智能体的平均报酬均能上升并趋于收敛。由于MA-DDQN算法适合处理离散变量,而本文研究场景中表征无人机移动的飞行距离和飞行角度属

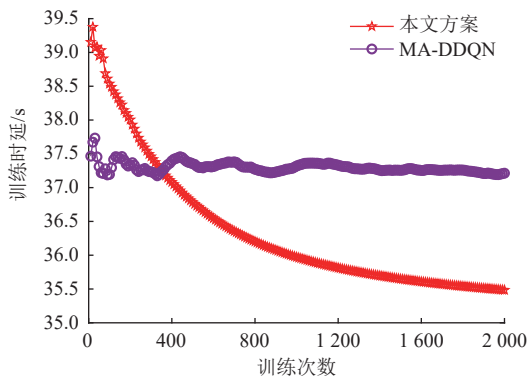


图3 不同算法下的联邦学习训练延迟对比

Fig.3 Comparison of federated learning training delay under different algorithm

于连续变量,因此,MA-DDQN算法需先对这些连续变量进行离散化处理后再执行迭代训练,而这限制了获得的无人机飞行距离和飞行角度的准确性。尽管MA-DDQN算法最终也能收敛至一个稳定区间,但其在报酬优化和收敛方面都逊于本文方案算法。这一结果充分证明了本文方案算法在提出的无人机数字孪生边缘网络中的优势。

图4对比了不同学习率下无人机智能体的平均报酬。从图中可以看出,随着训练次数增加,不同学习率下无人机智能体的平均报酬都能增加并收敛至稳定值,证明了提出的基于孪生网络模型的MA-DDPG优化算法可以通过探索不同的动作策略,如无人机 m 在第 t 时刻的飞行距离 $d_m(t)$,无人机飞行角度 $\varphi_m(t)$,子信道分配决策 $\phi_{n,m,k}(t)$,从而实现最大化无人机平均报酬的目标。此外,当Actor神经网络参数的学习率 $a = 10^{-4}$,Critic神经网络参数的学习率 $c = 10^{-4}$ 时,获得的无人机平均报酬最大。其次,当Actor神经网络参数的学习率 $a = 10^{-5}$,Critic神经网络参数的学习率 $c = 10^{-5}$ 时,获得的无人机平均报酬最小。这说明设置较大的学习率,可以获得较好的收敛性能。

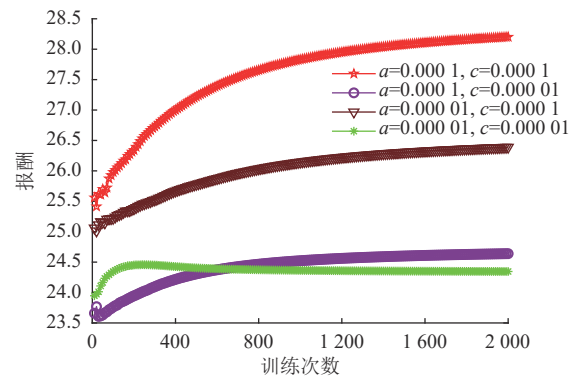


图4 不同学习率下的无人机平均报酬对比

Fig.4 Comparison of UAV average reward under different learning rates

图5展示了本文提出的方案与只优化信道和只优化轨迹的两种方案在联邦学习训练延迟方面的对比。根据报酬函数(见式(14)),联邦学习的训练延迟是计算报酬时的一个关键因素,其最小化是实现报酬最大化的必要条件,这与本文旨在最小化联邦学习训练延迟的目标相吻合。从图中可以观察到,随着训练迭代次数的增加,3种优化方案的训练延迟都有所下降,并最终趋于稳定。本文提出的方案在降低联邦学习训练延迟方面表现出了最佳的性能,这一优势源自于对无人机的飞行距离、飞行角度以及子信

道分配3个关键变量的综合优化。当且仅当这3个变量同时得到优化时,才能取得最佳效果。仿真结果证实了,相较于其他对照方案,本文提出的方案更有效地降低了联邦学习的训练延迟。

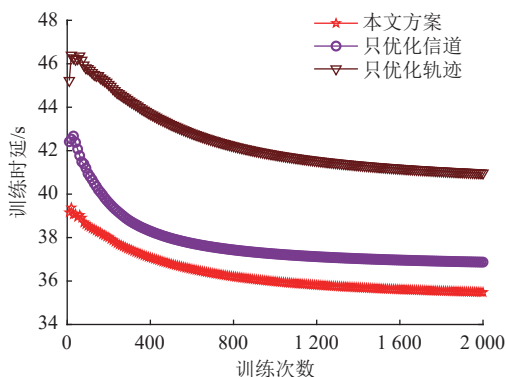


图5 不同优化方案下的联邦学习训练时延对比

Fig.5 Comparison of federated learning training delay under different optimization schemes

图6对比了3种优化方案实现的无人机数字孪生边缘网络联邦学习训练损失。从图中可以看出,随着联邦学习全局训练回合数增加,3种优化方案的联邦学习训练损失都能逐渐减小并趋于稳定值。其次,本文方案可以获得最小的联邦学习训练损失。这是因为提出的方案通过联合优化无人机飞行距离、无人机飞行角度和子信道分配3个优化变量,无人机可以聚合更多地面智能终端的本地模型参数,从而使全局模型参数训练损失快速降低。而在只优化子信道分配的优化方案中,无人机悬停于固定位置,无法采集服务区域内所有地面智能终端的本地模型参数,因而该方案的损失比本文方案大。在只优化无人机轨迹的优化方案中,子信道随机分配给地面智能终端,无法保证所有地面智能终端被分配足够子信道传输本地模型参数,因而该方案的损失最大。

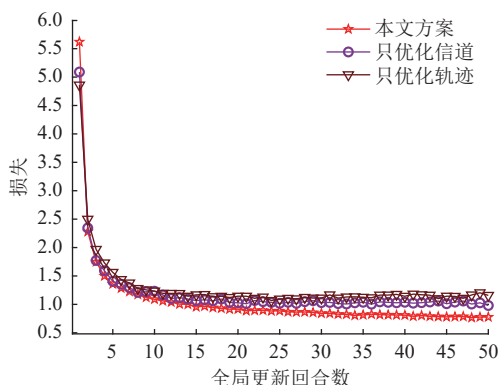


图6 不同优化方案下的联邦学习训练损失

Fig.6 Federated learning training losses under different optimization schemes

图7对比了3种优化方案实现的无人机数字孪生边缘网络联邦学习训练准确率。从图中可以看出,随着联邦学习全局训练回合数增加,3种优化方案的联邦学习训练准确率都能逐渐增加并趋于稳定值。其次,本文方案可以获得最大的全局模型准确率。与图6的分析一致,只优化子信道分配的方案中全局模型的准确率低于本文方案,而只优化无人机轨迹的方案中全局模型的准确率最低。

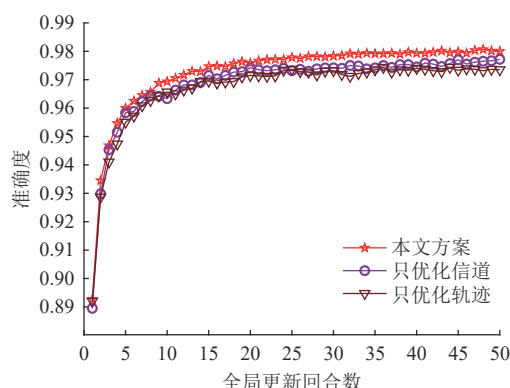


图7 不同优化方案下的联邦学习训练准确率

Fig.7 Federated learning training accuracy under different optimization schemes

4 结论

本文采用联邦学习,提出基于无人机数字孪生边缘网络的联邦学习资源调度机制。考虑无人机数字孪生边缘网络环境的动态时变性,建立联合无人机飞行距离、飞行角度以及无线网络子信道分配的联邦学习自适应资源优化模型,实现最小化无人机数字孪生边缘网络联邦学习时延的目标。提出基于孪生网络模型的MA-DDPG算法,求解联邦学习自适应资源优化模型。仿真结果表明,本文方案可显著降低联邦学习时延性能,同时保证联邦学习损失函数和准确率的优越性。下一步工作将研究无人机数字孪生边缘网络中集成联邦学习与现有隐私保护技术,例如差分隐私技术,并将设计新的隐私保护框架和资源优化机制。

参考文献:

- [1] GERACI G, RODRIGUEZ A, AZARI M, *et al.* What will the future of UAV cellular communications be? A flight from 5G to 6G [J]. *IEEE Communications Survey & Tutorials*, 2022, 24(3): 1304-1335.
- [2] JIANG L, CHEN B, XIE S L, *et al.* Incentivizing resource cooperation for blockchain empowered wireless power transfer in UAV networks [J]. *IEEE Transactions on Vehicu-*

- lar Technologies, 2020, 69(12): 15828-15841.
- [3] HAZRA K, SHAH V K, ROY S, *et al.* Exploring biological robustness for reliable multi-UAV networks [J]. *IEEE Transactions on Network and Service Management*, 2021, 18(3): 2776-2788.
- [4] TRIPATHY A, TRIPATHY K, MOHAPATRA G, *et al.* WeDoShare: a ridesharing framework in transportation cyber-physical system for sustainable mobility in smart cities [J]. *IEEE Consumer Electronics Magazine*, 2020, 9(4): 41-48.
- [5] HU C H, FAN W C, ZENG E, *et al.* Digital twin-assisted real-time traffic data prediction method for 5G-enabled Internet of Vehicles [J]. *IEEE Transactions on Industrial Informatics*, 2022, 18(4): 2811-2819.
- [6] WANG Z R, GUPTA R, HAN K, *et al.* Mobility digital twin: concept, architecture, case study, and future challenges [J]. *IEEE Internet of Things Journal*, 2022, 9(18): 17452-17467.
- [7] KONECNY J, MCMAHAN H B, RAMAGE D. Federated optimization: distributed machine learning for on-device intelligence[J]. arXiv: 1610.02527(2016-10-8) [2023-10-11]. <https://arxiv.org/abs/1610.02527>.
- [8] KHAN L U, SAAD W, HAN Z, *et al.* Federated learning for Internet of things: recent advances, taxonomy, and open challenges [J]. *IEEE Communications Surveys & Tutorials*, 2021, 23(3): 1759-1799.
- [9] 蒋丽, 谢胜利, 张彦. 面向6G网络的联邦学习资源协作激励机制设计[J]. *广东工业大学学报*, 2021, 38(6): 47-52,83. JIANG L, XIE S L, ZHANG Y. Incentivizing resource cooperation for federated learning in 6G networks [J]. *Journal of Guangdong University of Technology*, 2021, 38(6): 47-52,83.
- [10] SONG Y, WANG T, WU Y, *et al.* Non-orthogonal multiple access assisted federated learning for UAV swarms: an approach of latency minimization[C]//2021 International Wireless Communications and Mobile Computing (IWCMC). Harbin, China: IEEE, 2021: 1123-1128.
- [11] CHENG Z, XIA X, MIN M, *et al.* Auction-promoted trading for multiple federated learning services in UAV-aided networks [J]. *IEEE Transactions on Vehicular Technology*, 2022, 70(10): 10960-10974.
- [12] SUN W, XU N, WANG L, *et al.* Dynamic digital twin and federated learning with incentives for air-ground network [J]. *IEEE Transactions on Network Science and Engineering*, 2022, 9(1): 321-333.
- [13] LI B, LIU Y F, TAN L, *et al.* Digital twin assisted task offloading for aerial edge computing and networks [J]. *IEEE Transactions on Vehicular Technology*, 2022, 71(10): 10863-10877.
- [14] CHEN X, CHEN T, ZHAO Z, *et al.* Resource awareness in unmanned aerial vehicle-assisted mobile-edge computing system[C]//2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring). Antwerp, Belgium: IEEE, 2020.
- [15] RYAN L, YI W, AVIV T, *et al.* Multi-agent actor-critic for mixed cooperative- competitive environments[J]. arXiv: 1706.02275(2017-6-7) [2023-10-12]. <https://arxiv.org/abs/1706.02275>.
- [16] KRIZHEVSKY A, HINTON G. Learning multiple layers of features from tiny images[EB/OL]. (2009-04-08)[2023-12-28]. <http://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf>.
- (责任编辑: 张玮欣 英文审核: 熊荣斌)