

面向无蜂窝大规模 MIMO 的分布式智能内容缓存与 用户关联联合优化方法

李 云^{1,2}, 张承宇¹, 姚枝秀², 夏士超^{2*}, 谭 震³

(1. 重庆邮电大学通信与信息工程学院, 重庆 400065; 2. 重庆邮电大学计算机科学与技术学院, 重庆 400065;
3. 中国电子科技集团公司第十五研究所, 北京 100083)

摘 要: 在无蜂窝大规模 MIMO (Cell-Free massive Multiple-Input Multiple-Output, CF-mMIMO) 网络环境中, 业务需求差异化、环境高度动态化以及资源部署去中心化等特征, 制约了 CF-mMIMO 缓存部署和分发过程中多维网络资源的分配效率. 为此, 本文对去中心化 CF-mMIMO 场景中的多样化内容缓存和多用户关联问题展开研究. 首先, 基于 CF-mMIMO 场景中内容缓存与用户关联间的耦合关系, 研究并建立了内容缓存、用户关联和多维资源分配模型. 其次, 针对随机时变的网络环境和不完备的网络状态观测, 以最大化网络能效为目标, 将内容缓存、用户关联和资源分配问题抽象为分布式部分可观测马尔科夫决策过程. 而考虑到多样化内容缓存需求和广域差异化网络空间特征, 进一步提出一种基于图注意力网络的多智能体深度强化学习算法对内容缓存、用户关联和多维资源分配进行策略学习和优化. 最后, 仿真结果验证了所提算法在网络能效、系统吞吐量、缓存命中率方面具有明显的性能提升.

关键词: 无蜂窝大规模 MIMO; 内容缓存; 用户关联; 资源分配; 图注意力网络

基金项目: 国家自然科学基金 (No.62301099, No.62071077); 中国博士后科学基金 (No.2023MD734137)

中图分类号: TN929.5

文献标识码: A

文章编号: 0372-2112(2025)06-1932-11

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20240867

Joint Optimization of Distributed Intelligent Content Caching and User Association for CF-mMIMO

LI Yun^{1,2}, ZHANG Cheng-yu¹, YAO Zhi-xiu², XIA Shi-chao^{2*}, TAN Zhen³

(1. School of Communications and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China; 2. School of Computer and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China;
3. No.15 Institute of China Electronics Technology Group Corporation, Beijing 100083, China)

Abstract: In the cell-free massive MIMO (CF-mMIMO) networks, characterized by differentiated service requirements, highly dynamic conditions, and decentralized resource deployment, the efficiency of distributing multi-dimensional network resources during CF-mMIMO caching deployment is constrained. To address this, this paper conducts research on the problem of diverse content caching and multi-user association in decentralized CF-mMIMO scenarios. First, based on the coupling relationship between content caching and user association, models for content caching, user association, and multi-dimensional resource allocation are studied and established. Second, given the stochastic and time-varying network environment and incomplete network state observations, the content caching, user association, and resource allocation problem are abstracted as a distributed partially observable Markov decision process (POMDP) with the objective of maximizing network efficiency. Taking into account the diverse content caching requirements and wide spatial differentiation, a multi-agent deep reinforcement learning algorithm based on graph attention network is further proposed for strategic learning and optimization of content caching, user association, and multi-dimensional resource allocation. Finally, simulation results confirm that the proposed algorithm significantly enhances performance in terms of network efficiency, system throughput and cache hit rate.

Key words: cell-free massive MIMO; content caching; user association; resource allocation; graph attention networks

Foundation Item(s): National Natural Science Foundation of China (No.62301099, No.62071077); China Postdoctoral Science Foundation (No.2023MD734137)

1 引言

随着物联网和边缘计算技术的蓬勃发展,全球通信流量爆炸式增长,对无线网络架构提出了新的技术挑战^[1]. 由于传统蜂窝网络固有的结构化边界和中心化管理机制,已难以高效地应对当前复杂、高密度的通信业务需求. 为此,无线通信领域开始转向新型网络架构的研究和探索,其中 CF-mMIMO (Cell-Free massive Multiple-Input Multiple-Output)被认为是一个有潜力的解决方案. 不同于传统蜂窝 MIMO (Multiple-Input Multiple-Output)系统的固定边界,CF-mMIMO 通过大量分布式节点部署为用户提供服务,从而实现更为均匀的网络覆盖,并凭借其分布式的网络特征,为通信网络的灵活性和自适应性带来了显著提升^[2].

内容缓存作为一项提高无线网络资源利用率的新兴技术,能够根据用户业务的动态需求,在接入侧对传输的内容进行缓存部署,以提升无线链路的资源利用率,降低业务传输时延,并改善用户业务服务体验. CF-mMIMO 中广泛分布的天线节点及其去中心化的架构设计,为在网络边缘实施内容缓存提供了天然的优势.

内容传输时延和能效优化是 CF-mMIMO 内容缓存网络的研究热点之一. 为提高网络容量和系统能效,文献[3]提出了一种以用户为中心的 CF-mMIMO 内容缓存场景,并通过多接入点 (Access Point, AP) 间的协作内容缓存对服务内容传输时延和能耗进行了优化. 文献[4]以最大化能量效率为目标,并综合考虑了无线信道状态和用户业务请求,分别提出集中式和分布式的内容缓存部署方法. 为了提高系统的缓存命中率,文献[5]考虑了一种离线缓存辅助的 CF-mMIMO 场景,并提出了基于贪婪策略的自适应功率控制算法. 文献[6]研究了车联网中的协作缓存问题,针对内容新鲜度设计了一种基于内容年龄价值的最大增量算法,确保缓存内容的有效性.

内容部署与用户关联对 CF-mMIMO 网络性能具有重要影响. 文献[7]提出了一种多用户关联和资源分配联合优化方法,以提升系统吞吐量. 文献[8]研究了缓存空间受约束场景下的内容缓存和内容分发问题,并分别提出了基于用户请求感知的内容部署算法和基于缓存感知的用户关联算法. 用户请求内容的流行度是缓存部署决策重要依据. 文献[9]在内容流行度未知的网络场景中,基于历史内容请求数据,提出了一种基于聚类的长短期记忆网络对未来内容的请求趋势进行预测. 为了提高内容缓存命中率,文献[10]则提出一种基于马尔科夫链的内容请求预测模型,并基于该模型提出一种边缘缓存内容替换方法. 文献[11]研究了异构网络内容缓存场景中的

性能优化问题,并以最大化网络吞吐量为目标,分别提出了基于位置的用户关联策略和基于内容请求的用户关联策略. 文献[12]建立了一种网络功耗与用户服务质量联合优化模型,并从用户关联分组与功率资源分配两个角度进行了分解优化.

上述研究大多针对传统 CF-mMIMO 网络场景和一致性的用户业务需求,未充分考虑到多用户多内容 CF-mMIMO 网络环境中内容缓存部署、用户关联和资源分配间的耦合关系. 同时,在实际的网络场景中,用户业务需求多样化、资源部署特征去中心化以及网络环境高度动态化,也使得 CF-mMIMO 网络的资源管理问题变得十分复杂. 首先,用户业务需求类型和空间分布的多样化造成不同 AP 在内容缓存、资源分配等方面具有显著的空间差异性,并降低了 AP 节点间的协作效率. 其次,去中心化的资源部署特征制约了 AP 间的干扰管理性能和网络资源使用效率. 在去中心化的 CF-mMIMO 网络架构中,资源往往分布在多个地理位置分散的 AP 节点上,资源分布的不均匀性可能会导致不均衡的资源利用和较差的网络干扰协调水平^[13]. 因此,网络干扰协调和资源分配也需要充分考虑去中心化资源部署的空间特征,以优化各个 AP 节点的资源利用效率.

此外,用户设备倾向于连接到具有所需内容的 AP^[4],内容缓存策略直接影响用户关联决策. 同时,资源分配策略需要根据用户关联和可用的缓存内容来调整,以优化网络的整体性能. 不同网络元素间的耦合关系使得网络设计必须综合考虑所有这些因素,以实现最优的网络效率和用户满意度. 而 CF-mMIMO 网络中用户的频繁切换、网络状态的快速变化以及服务请求的时变特性等因素,造成网络状态和资源可用性具有深度不确定性特征. 内容缓存、用户关联以及资源分配需具备在非完全信息下实现高效策略制定能力.

针对上述问题,本文考虑到差异化用户需求、深度不确定的网络环境以及分布式的 CF-mMIMO 网络场景,提出一种基于图注意力网络 (Graph Attention network, GAT) 多智能体深度强化学习的内容缓存、用户关联和资源分配联合优化 (GAT based Multi-agent deep reinforcement learning for Joint Optimization of Content caching, user association and resource allocation in CF-mMIMO, GMJOC) 方法. 本文具体贡献如下:

(1) 针对去中心化 CF-mMIMO 网络场景中内容缓存、用户关联以及多维资源分配间的耦合关系,以网络能效最大化为目标,研究并建立了一种分布式 CF-mMIMO 场

景中的内容缓存、用户关联和资源分配联合优化模型。

(2) 针对动态时变的网络环境和不完备的网络状态观测, 将上述联合优化问题抽象为分布式部分可观测马尔科夫决策过程 (Decentralized Partially Observable Markov Decision Process, Dec-POMDP), 并设计了一种基于多智能体深度强化学习的分布式内容缓存和资源分配方法来自主决策内容缓存部署、用户关联和传输功率控制, 以处理因观测限制和环境不确定性带来的决策挑战。

(3) 考虑到 CF-mMIMO 场景中内容缓存需求的多样化及网络空间特征的广域差异化, 进一步采用图注意力网络来学习和捕捉网络空间特征。通过这种方法, 能够对内容下发过程中的干扰进行自适应控制, 并有效满足不同业务需求, 增强网络在复杂环境下的性能。

(4) 最后, 通过与现有研究的仿真对比分析, 验证了所提算法在网络能效、系统吞吐量、网络能耗和缓存命中率方面的性能。

2 系统模型

考虑一个多用户多内容的典型 CF-mMIMO 网络, 如图 1 所示, 该网络中包含 N 个 AP 和 M 个移动设备 (Mobile Device, MD), 分别定义 AP 和 MD 集合为 $\mathcal{N} = \{1, 2, \dots, N\}$ 和 $\mathcal{M} = \{1, 2, \dots, M\}$ 。其中, 不同 MD 具有差异化内容需求, 并根据当前网络状态关联到不同 AP; 不同 AP 根据服务范围内的内容需求缓存相应服务内容, 且通过光纤链路连接到中央处理单元 (Central Processing Unit, CPU)。网络运行在离散时隙, 定义时隙集合为 $\mathcal{T} = \{0, 1, 2, \dots\}$ 。

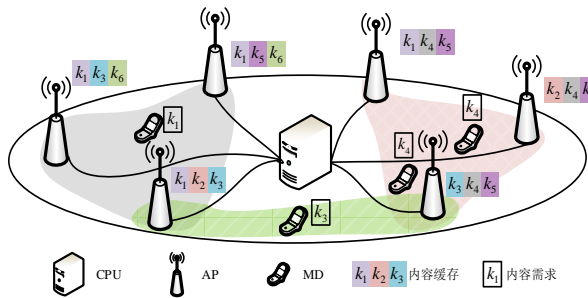


图 1 CF-mMIMO 网络 (不同 MD 关联不同 AP 集, 如不同的颜色区域; 不同 AP 缓存不同内容, 如不同颜色内容块。)

$$\begin{aligned}
 r_i(t) &= \sum_{j \in \mathcal{V}(t)} g_{ij}(t) x_j(t) + w_i(t) \\
 &= \underbrace{\sum_{j \in \mathcal{V}(t)} v_{ij}(t) g_{ij}(t) x_j(t)}_{\text{有用信号}} + \underbrace{\sum_{j \in \mathcal{V}(t)} v_{ij}(t) g_{ij}(t) x_j(t) + w_i(t)}_{\text{干扰信号}} \\
 &= \sum_{j \in \mathcal{V}(t)} v_{ij}(t) \sqrt{P_{ij}(t)} |g_{ij}(t)| q_i(t) + \sum_{j \in \mathcal{V}(t)} \sum_{i' \in \mathcal{B}(t), i' \neq i} v_{i'j}(t) \sqrt{P_{i'j}(t)} g_{ij}(t) g_{i'j}^*(t) q_i(t) + w_i(t)
 \end{aligned} \tag{4}$$

2.1 用户关联模型

在任一时隙, MD 基于与 AP 之间距离、信道条件以及内容部署状态等因素, 选择关联到不同的 AP。用 $\mathcal{V}_i(t)$ 表示在时隙 $t (t \in \mathcal{T})$ 与 MD $i (i \in \mathcal{M})$ 关联的 AP 集合, 用 $v_{ij}(t) (v_{ij}(t) \in \mathcal{V}_i(t))$ 表示 MD i 与 AP $j (j \in \mathcal{N})$ 之间的关联关系, 那么可以定义

$$v_{ij}(t) = \begin{cases} 1, & \text{MD } i \text{ 与 AP } j \text{ 关联} \\ 0, & \text{MD } i \text{ 不与 AP } j \text{ 关联} \end{cases} \tag{1}$$

为保证所有的 MD 均能被服务, 需满足 $0 < |\mathcal{V}_i(t)| \leq N$ 。此外, 并非所有 AP 都在为 MD 提供内容服务, 所有处于服务状态的 AP 集合为 $\mathcal{V}(t) = \bigcup_{i \in \mathcal{M}} \mathcal{V}_i(t)$, 且 $\mathcal{V}(t) \subseteq \mathcal{N}$ 。

2.2 信号模型

MD 向关联 AP 发起内容请求后, 若 AP 已缓存有该内容, AP 则经过下行无线信道将内容传输到相应 MD。AP j 到 MD i 的下行信道增益可以表示为

$$g_{ij}(t) = (d_{ij}/d_0)^{-\alpha} h_{ij}(t) \tag{2}$$

其中, d_{ij} 表示 AP j 和 MD i 之间的实际距离, d_0 表示参考距离, α 为路径衰减因子, $h_{ij}(t) \sim \mathcal{CN}(0, 1)$ 表示服从复高斯分布的小尺度衰落。

假设 AP j 到 MD i 传输的符号为 $q_i(t)$, 其中 $\mathbb{E}[|q_i(t)|^2] = 1$, $\mathbb{E}[q_i(t)] = 0$, 得到 AP j 具有完全信道状态信息的共轭波束形成传输信号为

$$x_j(t) = \sum_{i \in \mathcal{B}_j(t)} v_{ij}(t) \sqrt{P_{ij}(t)} \hat{g}_{ij}^*(t) q_i(t) \tag{3}$$

其中, $P_{ij}(t) (P_{ij}(t) \leq P_j^{\max})$ 是 AP j 分配给 MD i 的传输功率, P_j^{\max} 是 AP j 最大传输功率, $\hat{g}_{ij}^*(t)$ 为估计信道增益, $\mathcal{B}_j(t)$ 为被 AP j 服务的用户集合。

进一步可以得到 MD i 的接收信号模型如式 (4) 所示。其中, $\mathcal{V}_i^c(t) = \mathcal{V}(t) \setminus \mathcal{V}_i(t)$ 与 $\mathcal{B}(t) = \bigcup_{j \in \mathcal{N}} \mathcal{B}_j(t)$ 表示在时隙 t 所有具有服务需求的 MD 集合, $w_i(t)$ 是 MD i 接收到的噪声信号。根据式 (2) 和式 (4) 可知, 对于为不同用户提供服务的 AP, 其空间距离越近, $g_{ij}(t)$ 越大, 即产生的干扰越强。

根据式 (4), 可以得到 MD i 的接收信号速率如式 (5) 所示。

$$R_i(t) = \log_2 \left(1 + \frac{\left| \sum_{j \in \mathcal{V}(t)} \sqrt{P_{ij}(t)} |g_{ij}(t)| \right|^2}{\sum_{i' \in \mathcal{B}(t), i' \neq i} \left| \sum_{j \in \mathcal{V}(t)} \sqrt{P_{i'j}(t)} g_{ij}(t) g_{i'j}^*(t) \right|^2 + |w_i(t)|^2} \right) \quad (5)$$

进一步可以得到网络可达速率为

$$R_{\text{sum}}(t) = \sum_{i \in \mathcal{B}(t)} R_i(t) \quad (6)$$

2.3 服务请求模型

网络中有 F 种大小相等的服务内容, 定义 $\mathcal{F} = \{1, 2, \dots, F\}$ 表示所有内容的集合. AP 通常配备有限的存储资源, 仅能缓存部分内容, 而 CPU 具备充足的存储资源, 能够存储全部网络内容^[14, 15]. AP 需根据所关联的 MD 需求, 缓存不同的服务内容, 定义 F_j^{\max} 和 $\mathcal{F}_j(t)$ 分别为 APj 的缓存容量和 APj 在时隙 t 缓存的内容集合, 其中, $F_j^{\max} < F$ 且 $0 < |\mathcal{F}_j(t)| \leq F_j^{\max}$.

令 $f_i(t) \in \mathcal{F}$ 表示 MDi 请求的服务内容类型, 则 APj 接收到的内容请求集合为 $\mathcal{F}_j^{\text{req}}(t) = \bigcup_{i \in \mathcal{B}_j(t)} \{f_i(t)\}$.

根据 MDi 的内容偏好向量 \mathbf{g}_i 确定. 在实际的网络环境中, 不同区域通常具有不同的内容流行度, 即区域间的用户通常具有差异化的服务内容需求, 用户间的内容偏好向量通常不同. 定义 $\mathbf{g}_i = \{g_i^1, g_i^2, \dots, g_i^f, \dots, g_i^F\}$ 为 MDi 的内容偏好向量, g_i^f 为 f 的流行度排名. 本文中内容流行度建模为 Zipf 分布, 在时隙 t , 排名为 g_i^f 的内容被 MDi 请求的概率为

$$Z_i^f(t) = \frac{[g_i^f]^{-\beta_i}}{\sum_{j=1}^F [g_i^j]^{-\beta_i}}, \quad t \in \mathcal{T} \quad (7)$$

其中, $\beta_i \in [0, +\infty)$, $i \in \mathcal{M}$ 称为 MDi 的偏态因子. 当 β_i 逐渐增大时, 概率分布更加偏向于流行度高的内容.

2.4 能耗模型

若关联 AP 缓存了 MD 请求的内容, 则通过下行信道向 MD 发送该内容; 若关联 AP 未缓存, AP 将首先从 CPU 下载相应的服务内容, 再传输给 MD. 在 CF-mMIMO 内容缓存部署和分发过程中, 能耗主要包括三部分:

(1) AP 下行链路内容传输能耗

AP 为关联的 MD 提供服务内容, 不同的 AP 根据其功率传输策略将有不同的能耗, 在时隙 t , APj 的内容分发能耗定义为

$$P_j^{\text{AP-tr}}(t) = \sum_{i \in \mathcal{B}_j(t)} P_{ij}(t) \quad (8)$$

(2) AP 服务内容更新或替换的能耗

为适应 MD 内容请求的动态变化, AP 应根据需求,

缓存、更新或替换服务内容, APj 在时隙 t 缓存、更新或替换的服务内容集合为 $\mathcal{F}_j(t)/\mathcal{F}_j(t-1)$. 可以得到 APj 的内容更新或替换能耗为

$$P_j^{\text{AP-up}}(t) = P^{\text{FL}} \cdot \left| \mathcal{F}_j(t)/\mathcal{F}_j(t-1) \right| \quad (9)$$

其中, P^{FL} 为前传链路传输单位内容所需能耗.

(3) AP 服务内容转发能耗

若 AP 没有缓存 MD 请求的服务内容, AP 则需要从 CPU 处获取相应的内容并转发给 MD, MD 在时隙 t 需要经 APj 从 CPU 处获取的内容集合为 $\mathcal{F}_j^{\text{req}}(t)/\mathcal{F}_j(t)$, 则 APj 的缺失内容传输能耗为

$$P_j^{\text{AP-del}}(t) = P^{\text{FL}} \cdot \sigma^{\text{pen}} \cdot \left| \mathcal{F}_j^{\text{req}}(t)/\mathcal{F}_j(t) \right| \quad (10)$$

其中, σ^{pen} 表示从 AP 处获取一个未缓存内容的能耗因子.

综上, 可以得到 APj 在时隙 t 的能耗为

$$P_j(t) = P_j^{\text{AP-tr}}(t) + P_j^{\text{AP-up}}(t) + P_j^{\text{AP-del}}(t) \quad (11)$$

所有正在服务的 AP 总能耗为

$$P(t) = \sum_{j \in \mathcal{V}(t)} P_j(t) \quad (12)$$

2.5 问题建模

本小节以在一定的能耗下实现更高的网络速率作为目标, 将 MD 关联、AP 内容缓存和功率分配问题建模如下:

$$\text{P1: } \max_{\mathcal{F}(t), \mathcal{V}(t), \mathcal{P}(t)} EE = \frac{1}{T} \mathbb{E} \left\{ \sum_{t=0}^{T-1} \frac{R_{\text{sum}}(t)}{P(t)} \right\} \quad (13a)$$

$$\text{s.t. } 0 < |\mathcal{V}_i(t)| \leq N, \forall i, t \quad (13b)$$

$$v_{ij}(t) \in \{0, 1\}, \forall i \in \mathcal{M}, j \in \mathcal{N} \quad (13c)$$

$$P_{ij}(t) \leq P_j^{\max}, \forall i \in \mathcal{M}, j \in \mathcal{N} \quad (13d)$$

$$\sum_{i \in \mathcal{B}_j(t)} P_{ij}(t) \leq P_j^{\max} \quad (13e)$$

$$0 < |\mathcal{F}_j(t)| \leq F_j^{\max}, \forall j, t \quad (13f)$$

其中, 内容缓存策略为 $\mathcal{F}(t) = \{\mathcal{F}_1(t), \mathcal{F}_2(t), \dots, \mathcal{F}_M(t)\}$; MD 关联策略为 $\mathcal{V}(t) = \{\mathcal{V}_1(t), \mathcal{V}_2(t), \dots, \mathcal{V}_M(t)\}$; $\mathcal{P}(t) = \{\mathcal{P}_1(t), \mathcal{P}_2(t), \dots, \mathcal{P}_M(t)\}$ 表示功率分配策略, $\mathcal{P}_j(t) = \{P_{1j}(t), P_{2j}(t), \dots, P_{Mj}(t)\}$ 表示 APj 的功率分配集合. 约束条件式(13b)用于保证所有具有内容需求的 MD 均能被服务; 式(13c)用于描述 MD 与 AP 间的关联关系; 式(13d)和式(13e)表示任一 AP 为 MD 分配的传输功率不能超

过最大功率值;式(13f)表示 AP 在任一时隙缓存的内容不能超过 AP 的最大内容缓存容量.

3 基于图注意力多智能体强化学习的内容缓存、用户关联和资源分配联合优化算法

在实际的 CF-mMIMO 内容缓存网络中,用户自主式、随机接入会导致网络资源利用不均衡,严重影响资源的充分利用和用户服务体验. 在 CF-mMIMO 去中心化资源部署和高动态的网络环境中,难以实时获得精确的、全局的网络状态信息(如用户请求类型、信道状态信息、用户接入策略等),传统的集中式优化方法已难以适用. 此外,从优化问题 P1 可以看出,该问题是一个典型的混合整数非线性规划问题,由式(13b)~式(13f)可知,内容部署、用户关联和资源分配策略只与当前状态有关,而与过去的状态无关,是一个典型的马尔可夫决策问题. 因此,将上述优化问题进一步抽象为一个部分可观测的分布式 Dec-POMDP 问题,并进一步提出 GMJOC 方法.

3.1 问题转化

优化问题 P1 可以转化为有 N 个 AP 的 Dec-POMDP, 用元组 $\langle \mathcal{S}, \{\mathcal{O}_j\}_{j \in \mathcal{N}}, \{\mathcal{A}_j\}_{j \in \mathcal{N}}, R, \gamma \rangle$ 描述,其中,每个 AP 代表一个智能体, \mathcal{S} 表示 CF-mMIMO 全局网络环境状态, \mathcal{O}_j 是 APj 的局部观测空间, \mathcal{A}_j 是 APj 的动作空间, R 是奖励函数, $\gamma \in [0, 1)$ 表示折扣因子. 在时隙 t , 每个 AP 接收到局部观测 $\mathbf{o}_j(t) \in \mathcal{O}_j$, 并选取一个动作 $\mathbf{a}_j(t) \in \mathcal{A}_j$. 之后,可得所有智能体的联合动作,用 $\mathbf{a}(t) \in \mathcal{A} = \mathcal{A}_1 \times \mathcal{A}_2 \times \dots \times \mathcal{A}_N$ 表示. 智能体在执行联合动作后,环境返回一个全局奖励 $r(t) = R(\mathbf{s}(t), \mathbf{a}(t))$, 并将状态转移到下一状态 $\mathbf{s}(t+1)$, 其中, $\mathbf{s}(t) \in \mathcal{S}$ 为时隙 t 的环境状态. 接下来定义环境状态、局部观测、动作和奖励函数.

(1) 环境状态

在时隙 t , 环境状态包括所有 AP 的内容缓存、信道状态以及用户位置状态信息. 因此, $\mathbf{s}(t)$ 定义为

$$\mathbf{s}(t) = \{\mathcal{K}_1(t), \mathcal{K}_2(t), \dots, \mathcal{K}_N(t), \mathcal{G}_1(t), \mathcal{G}_2(t), \dots, \mathcal{G}_N(t), l_1(t), l_2(t), \dots, l_M(t)\} \quad (14)$$

其中, $\mathcal{K}_j(t) = \{k_{1j}(t), k_{2j}(t), \dots, k_{fj}(t), \dots, k_{Fj}(t)\}$ 为 APj 的内容缓存状态, $k_{fj}(t) \in \{0, 1\}$ 表示 APj 在时隙 t 是否缓存了内容 f , $k_{fj}(t) = 1$ 表示 APj 缓存了内容 f , $k_{fj}(t) = 0$ 表示 APj 未缓存内容 f ; $\mathcal{G}_j(t) = \{g_{1j}(t), g_{2j}(t), \dots, g_{ij}(t), \dots, g_{mj}(t)\}$ 表示 APj 与 MD 间的信道增益; $l_i(t)$ 表示 MDi 的位置信息.

(2) 局部观测

在部分可观测的 CF-mMIMO 环境中,任一 AP 仅能观测到当前时隙的内容缓存状态和网络中的用户位置信息. 因此, $\mathbf{o}_j(t)$ 表示为

$$\mathbf{o}_j(t) = \{\mathcal{K}_j(t), l_1(t), \dots, l_i(t), \dots, l_M(t)\} \quad (15)$$

(3) 动作

根据优化问题 P1, 优化变量是内容缓存、用户关联和功率分配. 则 APj 的动作 $\mathbf{a}_j(t)$ 表示为

$$\mathbf{a}_j(t) = \{\mathcal{F}_j(t), \mathcal{V}_j(t), \mathcal{P}_j(t)\} \quad (16)$$

(4) 奖励函数

所有智能体执行联合动作 $\mathbf{a}(t)$ 后, 环境将返回一个全局奖励 $r(t)$ 来评估联合动作. 根据优化问题 P1, 定义奖励函数为

$$r(t) = R(\mathbf{s}(t), \mathbf{a}(t)) = \frac{R_{\text{sum}}(t)}{P(t)} \quad (17)$$

在部分可观测的环境中, 智能体 APj 接收到局部观测值 $\mathbf{o}_j(t)$, 并根据其局部策略 π_j 选择动作 $\mathbf{a}_j(t)$. 定义 $\pi = \{\pi_j\}_{j \in \mathcal{N}}$ 表示所有智能体的联合策略, GMJOC 的目标是学习一种对内容缓存、用户关联和资源分配联合优化的策略, 以最大化衰减累计全局奖励 $\sum_{h \in T} \gamma^h r(t+h)$, 其中 γ 表示衰减因子. 因此, 联合动作价值函数可定义为

$$Q^\pi(\mathbf{s}(t), \mathbf{a}(t)) = \mathbb{E} \left[\sum_{h \in T} \gamma^h r(t+h) \mid \mathbf{s}(t), \mathbf{a}(t) \right] \quad (18)$$

其中, $\mathbb{E}[\cdot]$ 为数学期望运算, 动作值函数 $Q^\pi(\mathbf{s}(t), \mathbf{a}(t))$ 表示以 $\mathbf{s}(t)$ 为初始状态, π 为初始联合策略和 $\mathbf{a}(t)$ 为初始动作下的衰减累计全局奖励的期望, 最优联合策略 π^* 是使 $Q^\pi(\mathbf{s}(t), \mathbf{a}(t))$ 最大的联合策略.

3.2 GMJOC 算法

传统的 MARL 方法, 如独立 Q-learning, 对于求解 N 个 AP 的 Dec-POMDP 是可行的. 然而, 该方法存在以下挑战: (1) 由于智能体在学习过程中独立更新内容缓存、用户关联和资源分配策略, 使每个智能体训练的环境不平稳, 导致学习效率低下; (2) 无法有效利用智能体间潜在的空间相关性. 在 CF-mMIMO 网络场景中, 用户可获得的总传输速率与用户关联策略和区间干扰密切相关, 且距离越近的 AP 对彼此的干扰越大; 此外, 同一区域或距离相近的用户之间的内容需求越接近.

为了解决这些挑战, 并更好地利用用户业务需求和 AP 间信号干扰的潜在空间相关性, 将多智能体的环境构建为无向图 $G(\mathcal{V}, \mathcal{E})$, 其中 \mathcal{V} 是节点的集合, 每个节点代表一个智能体, \mathcal{E} 是表示节点之间连通性的边的集合. 本文采用基于图注意力的值分解网络, 将联合动作价值函数分解为局部动作价值函数基于注意力关系的

组合. 图注意力机制可以挖掘用户业务需求和 AP 之间信号干扰的空间相关性特征, 并计算每个智能体的局部动作值函数的权值. 本文采用集中式训练分布式执行 (Centralized Training with Decentralized Execution, CTDE) 的架构进行模型训练和策略学习. 如图 2 所示, 在集中式训练阶段, 每个 AP 首先接收环境的局部状态观测, 并执行动作, 可以得到所有智能体的联合动作, 执行联合动作后, 环境返回一个全局奖励, 并将状态转移到下一个状态. 而混合网络通过接收全局状态, 输出权重集合, 将所有智能体的 Q 函数混合后得到全局 Q 函数, 通过最小化损失函数进行梯度更新. 在执行阶段, 每个 AP 仅需根据其局部观测信息选择其内容缓存、用户关联和资源分配动作, 无需获取全局状态信息. 本文提出的 GMJOC 算法由 3 个独立的模块组成: (1) 每个智能体的局部动作价值网络; (2) 图注意力模块; (3) 联合动作价值混合模块.

(1) 局部动作价值网络. 每个智能体都配置了一个由多层感知器 (Multilayer Perceptron, MLP) 组成的深度 Q 网络 (Deep Q-Network, DQN), 如图 2 所示, 在时隙 t , 智能体 j 接收局部观测值 $\mathbf{o}_j(t)$, 并输出一个局部动作价值函数 $Q_j(\mathbf{o}_j(t), \mathbf{a}_j(t))$.

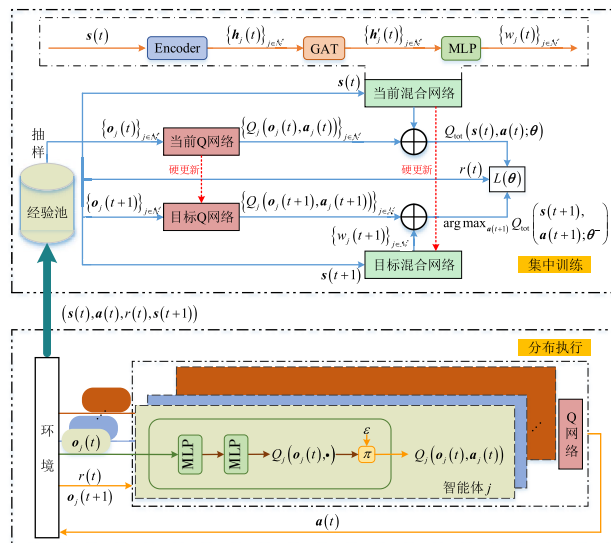


图 2 GMJOC 算法框架

(2) 图注意力模块. 如图 2 所示, 环境状态 $\mathbf{s}(t)$ 首先输入一个编码器, 将 $\mathbf{s}(t)$ 编码为局部潜在表示向量 $\mathbf{h}_1(t), \mathbf{h}_2(t), \dots, \mathbf{h}_N(t)$, 其中, $\mathbf{h}_j(t)$ 为每个节点的特征表示. 然后, 采用 GAT 自适应地捕获智能体之间的相关性.

在 CF-mMIMO 中, 任一节点有一个由边集 \mathcal{E} 确定的邻居节点集合 $\mathcal{H}_j (\mathcal{H}_j \subseteq \mathcal{N})$, 如果节点之间有边, 则为邻居. 在 GAT 中, 节点 j 与其相邻节点 $j' (j' \in \mathcal{H}_j)$ 之间的注意力系

数可由 $e_{jj'} = \text{att}(\mathbf{W} \cdot \mathbf{h}_j(t), \mathbf{W} \cdot \mathbf{h}_{j'}(t))$ 计算得到, 其中 $\text{att}(\cdot)$ 为自注意力机制, \mathbf{W} 为对应的可学习的权重矩阵. 注意系数 $e_{jj'}$ 表示邻居节点 j' 的特征对节点 j 的重要性. 在本文中, 只考虑节点 j 的一阶邻居节点 (包括节点 j). 为了便于在不同节点间进行比较, 使用 softmax 函数对 $e_{jj'}$ 进行归一化, 得到归一化注意系数:

$$\alpha_{jj'} = \text{softmax}(e_{jj'}) = \frac{\exp(e_{jj'})}{\sum_{h \in \mathcal{H}_j} \exp(e_{jh})} \quad (19)$$

为了稳定学习过程, 本文采用多头注意力机制. 在得到归一化注意系数后, 具有 L 个独立注意机制的节点 j 的输出特征表示向量可由下式给出:

$$\mathbf{h}_j'(t) = \parallel_{l=1}^L \sigma \left(\sum_{j' \in \mathcal{H}_j} \alpha_{jj'}^l \mathbf{W}^l \mathbf{h}_{j'}(t) \right) \quad (20)$$

其中, σ 为非线性函数, \parallel 表示拼接操作, l 表示注意力机制的序号, $\alpha_{jj'}^l$ 表示在第 l 个注意力头节点 j 对 j' 的权重, \mathbf{W}^l 表示第 l 个注意力头的权重矩阵. 然后, MLP 网络以 $\mathbf{h}_j'(t)$ 作为输入, 并为智能体 j 的局部动作价值函数生成权重 $w_j(t)$.

(3) 混合模块. 根据上述分析, 可得混合模块的图注意力权重 $\{w_j(t)\}_{j \in \mathcal{N}}$. 则联合动作价值函数可表示为

$$Q_{\text{tot}}(\mathbf{s}(t), \mathbf{a}(t)) = \sum_{j \in \mathcal{N}} w_j(t) Q_j(\mathbf{o}_j(t), \mathbf{a}_j(t)) \quad (21)$$

最后, 通过最小化损失函数的方式进行模型训练, 即

$$L(\theta) = \sum_{x=1}^X (y_{\text{tot}}^x - Q_{\text{tot}}(\mathbf{s}, \mathbf{a}; \theta))^2 \quad (22)$$

其中, $y_{\text{tot}} = r + \gamma \max_{\mathbf{a}'} Q_{\text{tot}}(\mathbf{s}', \mathbf{a}'; \theta^-)$, θ 和 θ^- 分别为评估和目标网络参数, X 表示从经验回放池中随机采样小批量样本数, x 表示样本序号.

4 仿真与分析

本节对 GMJOC 算法进行仿真验证. 首先给出了 CF-mMIMO 缓存网络环境下的仿真参数, 然后将 GMJOC 算法与其他算法进行性能对比和分析.

(1) 基于多智能体深度强化学习的 CF-mMIMO 内容缓存、用户关联和资源分配联合优化 (Multi-agent deep reinforcement learning based Joint Optimization of Content caching, user association and resource allocation in CF-mMIMO, MJOC) 算法, MJOC 算法未采用图注意力层.

(2) 基于多智能体深度确定性策略梯度的分布式用户关联与缓存 (Distributed User Association and Content caching, DUAC) 算法^[4]. 训练过程中每个智能体的

观测包含自身的状态信息以及其他智能体的缓存状态和位置状态信息。

(3)基于DDPG的用户关联和资源管理(User Association and Resource Management, UARM)算法^[16]。训练过程中每个智能体仅能观测到自身的缓存状态和位置状态,无法获取其他智能体的缓存状态和位置状态进行学习。

(4)随机缓存(Random Cache, RC)算法。AP随机做出内容缓存、用户关联和功率分配联合动作。

4.1 仿真参数设置

在 $1\text{ km} \times 1\text{ km}$ 范围内,考虑一个包含6个AP和4个MD的CF-mMIMO缓存网络。根据文献[4,5],系统带宽设置为40 MHz,每个AP的最大下行传输功率为100 mW,路径衰减因子 α 设为2,高斯信道噪声功率设为 $1 \times 10^{-13}\text{ W}$ 。假设网络中总共有[50,100]种服务,AP缓存容量设为30。由于不同区域用户具有不同内容需求,相邻区域用户具有相似的内容需求,定义MD1和MD2的内容偏好向量相近,MD3与MD4的内容偏好向量相近。每个用户对不同种类服务的偏好不同,即 $\beta_1 = 1.2, \beta_2 = 1.5, \beta_3 = 1.8, \beta_4 = 2.0$ 。

用户 i 在时刻 t 的命中结果用 $H_i(t)$ 表示,定义在时刻 t 所有为用户 i 服务的AP中已缓存的请求文件,即 $f_i(t) \in \mathcal{F}_m(t), \forall m \in \mathcal{V}_i(t)$ 。缓存命中率定义为

$$H(t) = \frac{1}{|\mathcal{B}(t)|} \sum_{i \in \mathcal{B}(t)} 1_{H_i(t)} \quad (23)$$

其中, $1_{H_i(t)}$ 是示性函数,如果事件发生,函数值等于1,否则为0。

对于GMJOC算法,每个智能体的神经网络为2层MLP网络,神经元个数分别为128和64,GAT模块中MLP层神经元个数为64,学习率为0.0001,折扣因子 γ 为0.98,Q网络每隔100步更新一次,经验池大小设置为50000,每个轮次的步长为100,在每一步迭代中,智能体采用贪心策略。

GMJOC算法中智能体的奖励定义为每个轮次中所有迭代次数的平均值,MJOC、DUAC和UARM算法的参数与GMJOC算法保持一致。

4.2 仿真分析

本文将GMJOC算法与MJOC、DUAC、UARM以及RC算法进行对比分析。首先对比了不同算法的收敛性能,然后在网络能效、吞吐量、能耗以及缓存命中率4个方面进行性能比较。文中能效与智能体获得的奖励等价。

(1)算法收敛性能对比

图3为GMJOC、MJOC、DUAC、UARM和RC算法的平均奖励随训练轮次增大的变化曲线。随着训练轮次的增加,智能体开始积累经验,不断优化自身的

内容缓存、用户关联和功率分配策略。GMJOC算法最终在400轮左右趋于稳定,奖励值约为4。GMJOC算法由于采用了GAT机制,通过捕捉用户业务需求和AP信号干扰的空间特征,能够学习到更优的策略,从而实现更优网络能效。MJOC算法对全局状态进行简单的编码处理,未能关注到不同AP对于全局奖励的重要性,无法利用全局价值函数得到更优的联合策略,性能不及GMJOC算法。DUAC算法虽然也利用其他智能体的状态信息进行训练,但是未能进一步对这些状态信息进行特征学习,性能次于GMJOC算法。UARM算法中每个智能体仅根据自身的观测和动作经验进行学习,智能体之间不进行信息共享,无法获取全局状态信息,因而收敛较慢,无法得出较优的内容缓存、用户关联和功率分配联合策略。RC算法无法对过去的经验进行学习,性能最差。

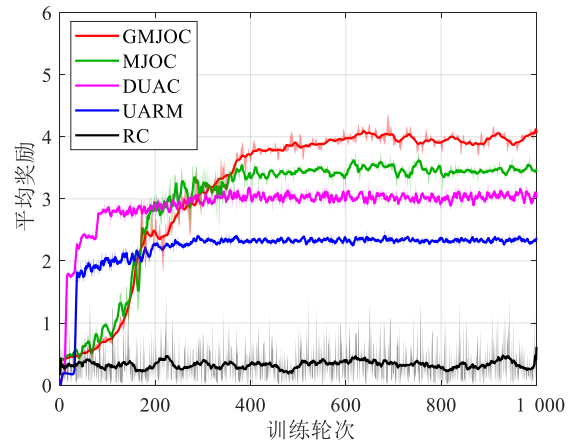


图3 算法收敛性能

(2)不同AP最大传输功率下的性能对比

图4分别对比了GMJOC、MJOC、DUAC、UARM和RC算法在不同AP最大传输功率下的能效、吞吐量、能耗和缓存命中率。随着AP最大传输功率的增加,AP能够给用户分配更多的功率,网络吞吐量逐渐增大,系统能耗随之缓慢增加,但能耗增速比网络速率慢,因此能效逐渐增大。而系统可达速率是对数函数,当AP最大传输功率达到800 mW时,该值趋于平稳。而缓存命中率随着AP最大传输功率的增加几乎没有变化,传输功率的增大对系统吞吐量和能耗的影响更加明显。所提GMJOC算法的性能较其他4种算法更优,因为GMJOC算法采用的注意力机制能够根据不同的权重整合来自不同AP的内容缓存和地理位置信息,进而学习到各AP之间的空间相关性,引导智能体输出更好的策略。

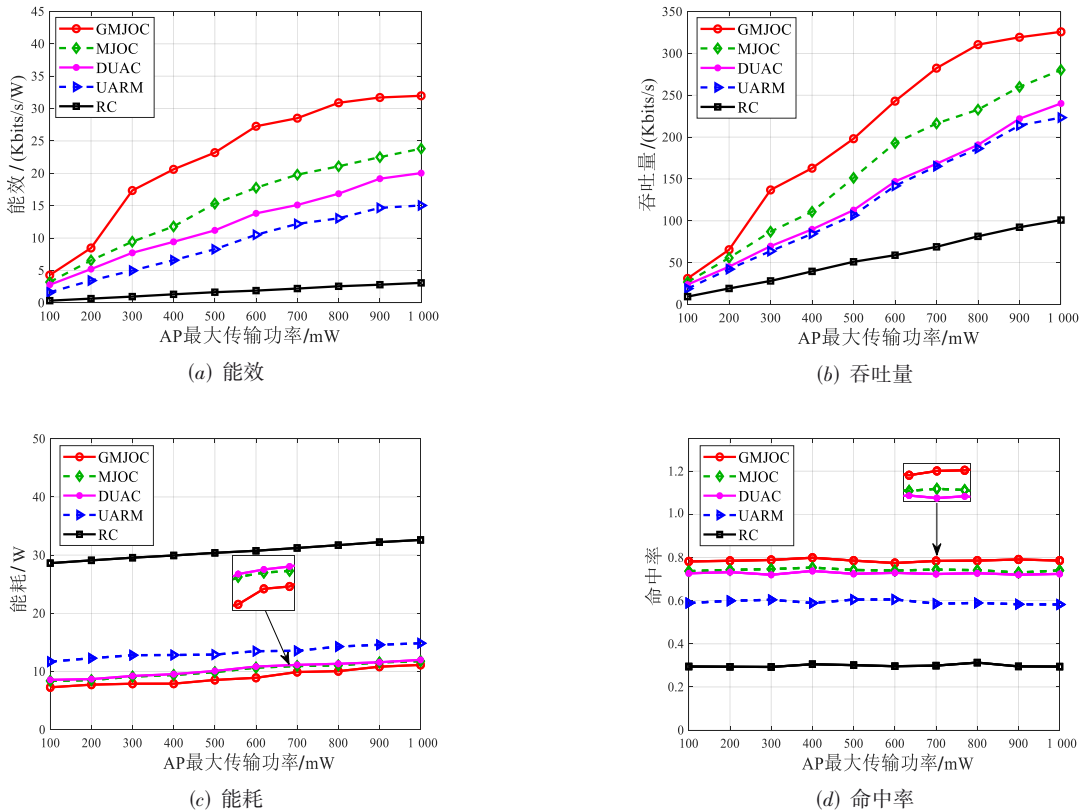


图 4 不同 AP 最大传输功率下的性能对比

(3) 不同 AP 缓存容量下的性能对比

图 5 分别对比了 GMJOC、MJOC、DUAC、UARM 和 RC 算法在不同 AP 缓存容量下的能效、吞吐量、能耗和缓存命中率。根据实验结果,能效和缓存命中率随着 AP 缓存容量的增加而逐渐变大,这是由于 AP 缓存空间的增加使得 AP 在同一时间能够缓存更多的服务,因此系统的缓存命中率逐渐增大,用户的请求更多地通过 AP 满足,这意味着前传链路的访问频率在减小,相应的内容转发能耗也在减小。而网络吞吐量随着 AP 缓存容量增加变化不明显,AP 缓存容量的变化没有同时影响所有指标的性能,因此系统能效逐渐增大。由于用户的内容请求概率服从 Zipf 分布,请求排名靠后的服务的概率相对较低,即这部分服务对于缓存命中率的提升贡献较小,系统的缓存命中率并不会线性增大,因此在 AP 缓存容量较小时,增大缓存容量对缓存命中率的提升影响较为明显,当 AP 缓存容量增大至 40 时,缓存命中率的增幅较小。GMJOC 算法的性能优于其他 4 种算法,证明了 GMJOC 算法的优势。

(4) 不同总内容数下的性能对比

图 6 分别对比了 GMJOC、MJOC、DUAC、UARM 和 RC 算法在不同网络总内容数下的能效、吞吐量、能

耗和缓存命中率。在 AP 缓存容量不变的情况下,随着系统总内容数的增加,系统吞吐量和缓存命中率均逐渐减小,系统能耗随之增大,这是因为总内容数的增加直接导致智能体内容缓存动作空间增大,对智能体来说任务的学习难度增大,内容缓存的不确定性将导致前传链路上传输更多的内容,最终导致能耗增加。所有算法中 GMJOC 算法性能最佳,结合 GAT 机制,GMJOC 算法能够在聚合节点信息时赋予不同的权值,有效提取不同区域内的内容流行度、信道状态等信息的特征,学习到 AP 间的空间相关性,智能体据此优化自身的策略。MJOC 算法缺少图注意力层,在输入全局状态后,智能体无法关注环境中其他智能体的信息,因此奖励比 GMJOC 算法小。另外,DUAC 算法中智能体虽然能够观测到其他智能体的状态信息,但未能对获取到的信息进行进一步的学习,因此性能不及 GMJOC 算法。UARM 算法中智能体之间没有交互,无法获取其他智能体的观测信息,整个网络环境的不确定性增加,因此性能更差。

(5) 不同内容流行度下的性能对比

图 7 在不同的内容流行度参数下对比了 GMJOC、MJOC、DUAC、UARM 和 RC 算法的能效、吞吐量、能耗

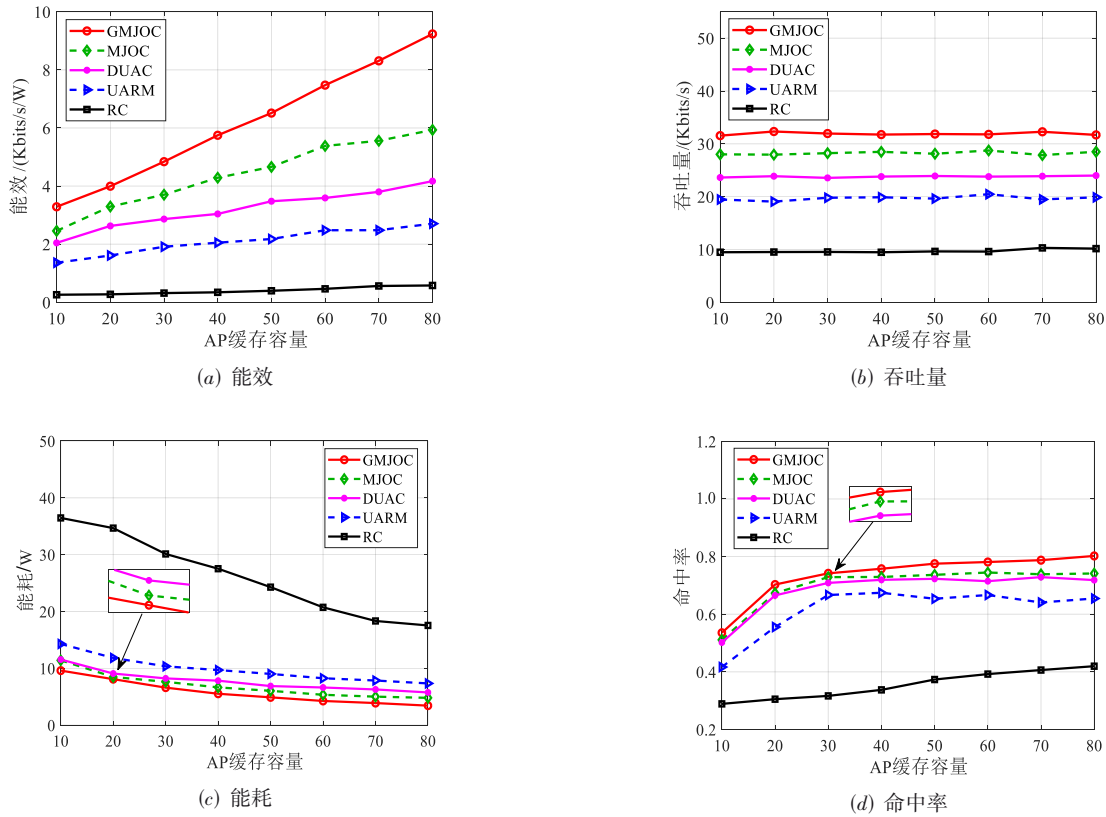


图5 不同AP缓存容量下的性能对比

和缓存命中率. 从图7可以看出,随着 β 的增大,缓存命中率逐渐增大,用户更多地从AP处获取内容,能耗逐渐减小,而系统吞吐量基本不变,因此系统能效逐渐增大. β 较小时,因为用户请求分散在大量内容上,如果要满足用户请求,则需要向CPU获取缺失的内容文件,这会增加前传链路上的传输能耗. β 的增大会使用户的请求概率分布集中于排名靠前的内容,即不同用户的请求具有更加明显的特征,智能体在进行缓存决策时更加倾向于缓存这部分内容.除RC算法外,其余4种算法的能效逐渐增加,说明智能体能够学习到 β 的变化.与其他算法相比,GMJOC算法的性能最好,在训练过程中,图注意力网络能够对节点的信息进行差异化处理,从而输出更有利于不同节点的缓存决策,提高系统的缓存命中率.而MJOC和DUAC算法中智能体之间虽然共观测信息,但没有对节点的输入数据进行差异化处理,性能比GMJOC算法差,UARM算法每个节点仅根据有限的本地信息做出决策,在4种智能算法中性能最差.

5 结束语

本文针对CF-mMIMO内容缓存网络场景中差异化用户内容需求、深度不确定网络环境以及去中心化资源部署特征,提出了一种基于GAT多智能体深度强化学习的内容缓存、用户关联和资源分配联合优化方法.首先,考虑到内容缓存、用户关联以及多维资源分配间的耦合关系,以及CF-mMIMO动态时变的网络环境和不完善的网络状态观测特征,研究并建立了内容缓存、用户关联和资源分配联合优化模型,并将联合优化问题抽象为分布式部分可观测马尔科夫决策过程,设计了一种基于多智能体深度强化学习的分布式内容缓存和资源分配方法来自主决策内容缓存部署、用户关联和传输功率控制.而考虑到CF-mMIMO场景中多样化内容缓存需求和广域差异化网络空间特征,采用GAT学习和捕捉网络空间特征,以实现内容下发过程中自适应干扰控制,并满足不同业务需求.最后,通过仿真验证了所提算法在网络能效、系统吞吐量、缓存命中率方面的性能.

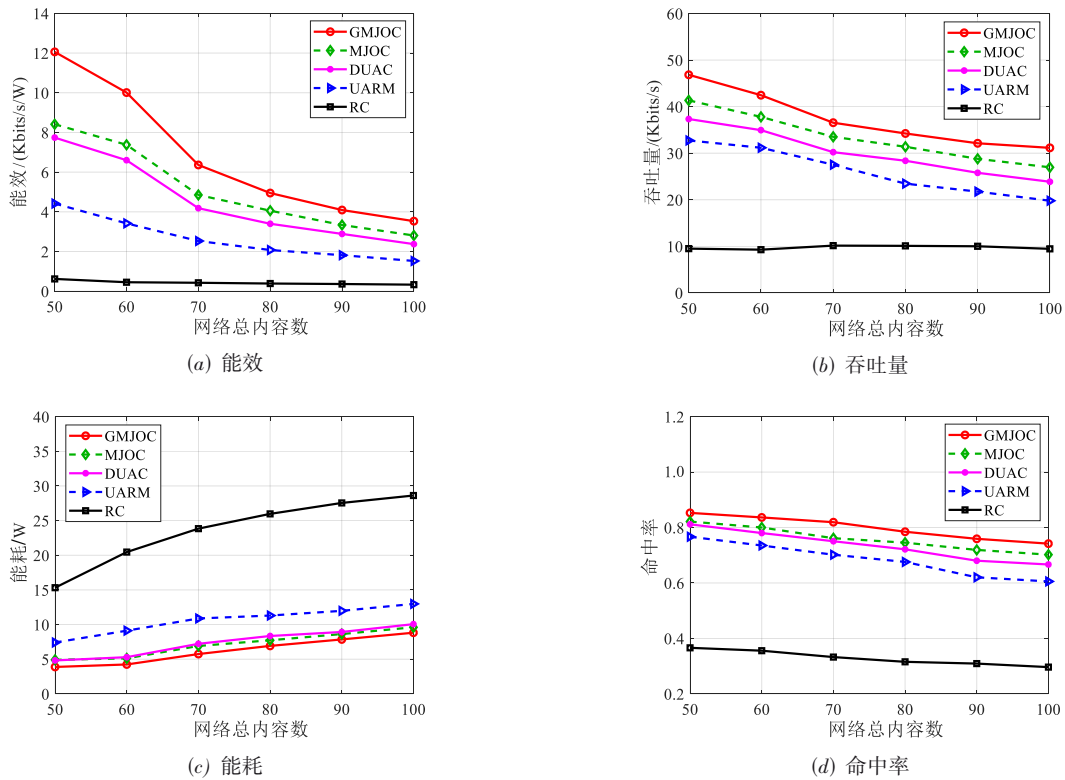


图 6 不同网络总内容数下的性能对比

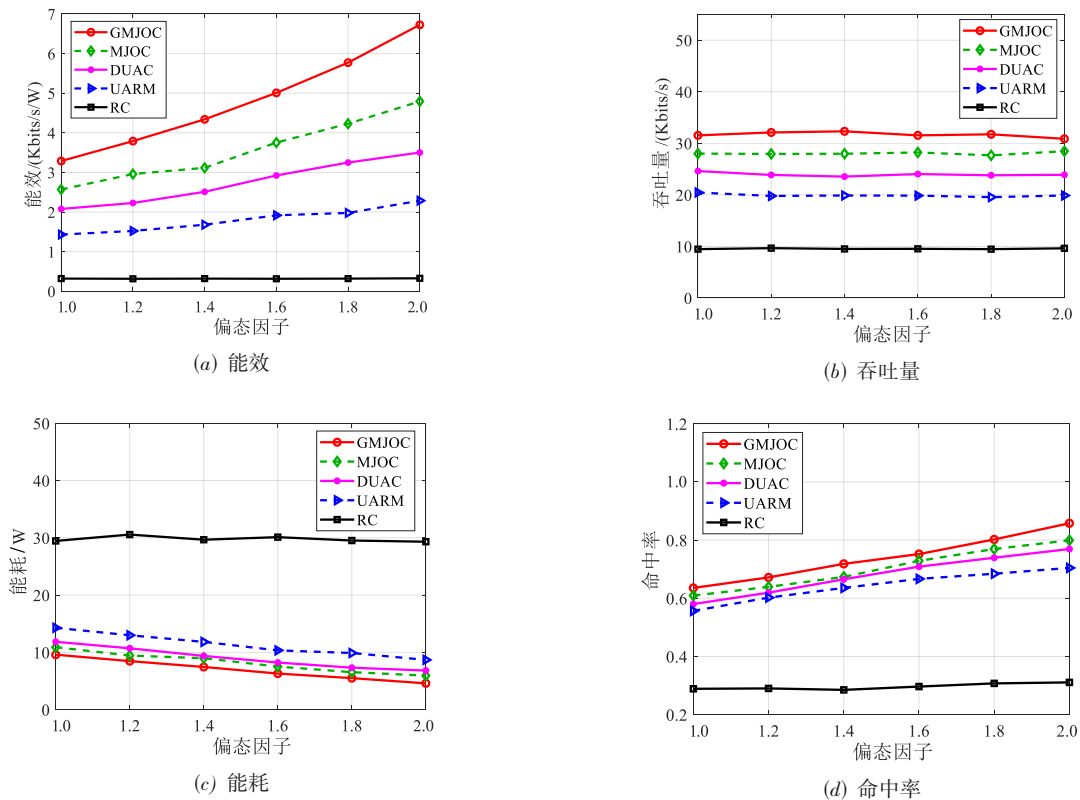
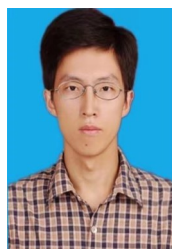


图 7 不同内容流行度参数下的性能对比

参考文献

- [1] XIA S, YAO Z, LI Y, et al. Distributed computing and networking coordination for task offloading under uncertainties[J]. IEEE Transactions on Mobile Computing, 2024, 23(5): 5280-5294.
- [2] AMMAR H A, ADVE R, SHAHBAZPANAH S, et al. User-centric cell-free massive MIMO networks: A survey of opportunities, challenges and solutions[J]. IEEE Communications Surveys & Tutorials, 2022, 24(1): 611-652.
- [3] RANASINGHE V, RAJATHEVA N, LATVA-AHO M. Graph neural network based access point selection for cell-free massive MIMO systems[C]//2021 IEEE Global Communications Conference (GLOBECOM). Piscataway: IEEE, 2021:1-6.
- [4] CHUANG Y C, CHIU W Y, CHANG R Y, et al. Deep reinforcement learning for energy efficiency maximization in cache-enabled cell-free massive MIMO networks: Single- and multi-agent approaches[J]. IEEE Transactions on Vehicular Technology, 2023, 72(8): 10826-10839.
- [5] CHEN S, ZHANG J, BJÖRNSSON E, et al. Wireless caching: Cell-free versus small cells[C]//ICC 2021 - IEEE International Conference on Communications. Piscataway: IEEE, 2021: 1-6.
- [6] 林志坚, 侯映, 曹晓晓, 等. 车联网中基于信息年龄价值的边缘缓存策略[J]. 电子学报, 2023, 51(12): 3410-3421.
- LIN Z J, HOU Y, CAO X X, et al. Edge caching scheme based on value of information age in the internet of vehicles [J]. Acta Electronica Sinica, 2023, 51(12): 3410-3421. (in Chinese)
- [7] HUANG X, ZHAO S, GAO X, et al. Online user-AP association with predictive scheduling in wireless caching networks[J]. IEEE Transactions on Mobile Computing, 2022, 21(6): 2116-2129.
- [8] WANG R, SHEN M, HE Y, et al. Joint access points-user association and caching placement strategy for cell-free massive MIMO systems based on soft actor-critic algorithm[J]. IEEE Communications Letters, 2024, 28(2): 347-351.
- [9] LI D, DING H, ZHANG H, et al. Deep learning-enabled joint edge content caching and power allocation strategy in wireless networks[J]. IEEE Transactions on Vehicular Technology, 2024, 73(3): 3639-3651.
- [10] WANG Z, HU J, MIN G, et al. Agile cache replacement in edge computing via offline-online deep reinforcement learning[J]. IEEE Transactions on Parallel and Distributed Systems, 2024, 35(4): 663-674.
- [11] TAN X, WANG S, JI L, et al. Hybrid-coding based content access control for information-centric networking[J]. IEEE Transactions on Wireless Communications, 2024, 23(7): 6765-6777.
- [12] 丁青锋, 李怡浩, 徐梦引. 去蜂窝大规模MIMO-NOMA系统能效优化算法[J]. 电子学报, 2023, 51(8): 2020-2029.
- DING Q F, LI Y H, XU M Y. Energy efficiency optimization algorithm for cell-free massive MIMO-NOMA systems[J]. Acta Electronica Sinica, 2023, 51(8): 2020-2029. (in Chinese)
- [13] YAO Z X, XIA S C, LI Y, et al. Cooperative task offloading and service caching for digital twin edge networks: A graph attention multi-agent reinforcement learning approach[J]. IEEE Journal on Selected Areas in Communications, 2023, 41(11): 3401-3413.
- [14] HU Z, ZHONG R, FANG C, et al. Caching-at-STARS: The next generation edge caching[J]. IEEE Transactions on Wireless Communications, 2024, 23(8): 8372-8387.
- [15] 夏士超, 张承宇, 姚枝秀, 等. 一种无蜂窝大规模MIMO系统的策略优化算法: CN202411149227.2[P]. 2025-07-15.
- [16] CHOU P Y, CHEN W Y, WANG C Y, et al. Pricing-based deep reinforcement learning for live video streaming with joint user association and resource management in mobile edge computing[J]. IEEE Transactions on Wireless Communications, 2022, 21(6): 4310-4324.

作者简介



李云 男, 1974年5月出生于四川省南充市. 现为重庆邮电大学教授、博士生导师. 主要方向为云计算、边缘智能、无线网络资源管理等.

E-mail: liyun@cqupt.edu.cn



夏士超 男, 1991年1月出生于山东省德州市. 现为重庆邮电大学讲师、硕士生导师. 主要研究方向为边缘智能、通感算一体化、无线网络资源管理等.

E-mail: xiashichao@cqupt.edu.cn