基于强化学习的非正交多址接入和移动边缘计算联合系统信息年龄更新

李保罡[①] 石 泰*[①] 陈 静[②] 李诗璐[①] 王 宇[①] 张天魁[③]

[①](华北电力大学 保定 071003)

[②](国网河南省电力公司许昌供电公司 许昌 461000)

[③](北京邮电大学 北京 100876)

摘 要：物联网发展对信息时效性的需求越来越高，信息新鲜度变得至关重要。为了维持信息新鲜度，在非正交多址接入(NOMA)和移动边缘计算(MEC)的联合系统中，对多设备单边缘计算服务器的传输场景进行了研究。在该场景中，如何分配卸载任务量和卸载功率以最小化平均更新代价是一个具有挑战性的问题。该文考虑到现实中的信道状态变化情况，基于多代理深度确定性策略梯度(MADDPG)算法，考虑信息新鲜度影响，建立了最小化平均更新代价的优化问题，提出一种寻找最优的卸载因子和卸载功率决策。仿真结果表明，采用部分卸载的方式可以有效地降低平均更新代价，利用MADDPG算法可以进一步优化卸载功率，经比较，MADDPG算法在降低平均更新代价方面优于其他方案，并且适当地减少设备数量在降低平均更新代价方面效果更好。

关键词：非正交多址接入；移动边缘计算；信息年龄；多代理深度确定性策略梯度

中图分类号：TN91 **文献标识码**：A **文章编号**：1009-5896(2022)12-4238-08

DOI: 10.11999/JEIT211021

Age of Information Updates in Non-Orthogonal Multiple Access-mobile Edge Computing System Based on Reinforcement Learning

LI Baogang[①] SHI Tai[①] CHEN Jing[②] LI Shilu[①]

WANG Yu[①] ZHANG Tiankui[③]

[①](*North China Electric Power University*, *Baoding* 071003, *China*)

[②](*Xuchang Power Supply Company of State Grid Henan Electric Power Company*, *Xuchang* 461000, *China*)

[③](*Beijing University of Posts and Telecommunications*, *Beijing* 100876, *China*)

Abstract: With the development of the Internet of Things, the demand for timeliness of information is increasing, and the freshness of information is becoming crucial. In order to maintain the freshness of information, the transmission scenario of multiple devices and single Mobile Edge Computing (MEC) server is studied in the joint system of Non-Orthogonal Multiple Access (NOMA) and MEC. In this scenario, how to allocate the amount of unload tasks and unload power to minimize the average update cost is a challenging problem. Considering the channel state variation in reality, an optimal unloading factor and unloading power decision are proposed based on Multi-Agent Deep Deterministic Policy Gradient (MADDPG) algorithm. Simulation results show that partial unloading can effectively reduce the average update cost, and MADDPG algorithm can further optimize the unloading power. By comparison, MADDPG algorithm is better than other schemes in reducing the average update cost, and the appropriate reduction of the number of equipment is better in reducing the average update cost.

Key words: Non-Orthogonal Multiple Access(NOMA); Mobile Edge Computing (MEC); Age of Information (AoI); Multi-Agent Deep Deterministic Policy Gradient(MADDPG)

收稿日期：2021-09-26；改回日期：2021-12-16；网络出版：2022-01-13

*通信作者： 石泰 taishist@163.com

基金项目：国家自然科学基金 (61971190)，中央高校基本科研业务费专项资金项目(2019MS089)，北京市自然科学基金(4164101)，河北省高等学校科学技术研究项目(ZD2021406)

Foundation Items: The National Natural Science Foundation of China (61971190), The Fundamental Research Funds For the Central Universities (2019MS089), The Natural Science Foundation of Beijing (4164101), The Science and Technology Research Project of Colleges and Universities in Hebei (ZD2021406)

1　引言

随着车载网络、虚拟现实等实时应用的发展，信息年龄(Age of Information, AoI)成为衡量信息新鲜度的一个重要标准。它被定义为目的端接收的最新更新信息自产生后所经过的时间[1]。由于在智能工厂、智慧型监控等创新应用中，终端设备不再是简单的数据收集，而是经过数据处理才能显现出所需的状态信息，因此，在AoI中引入计算受到了人们的广泛关注[2]。考虑到终端设备有限的电池容量和计算资源，移动边缘计算(Mobile Edge Computing, MEC)被认为是一种处理终端设备计算问题的有效解决方法[3]。在当前物联网的场景中，例如无人驾驶、超清视频和增强现实等，MEC可以满足这些任务的高计算要求，而非正交多址(Non-Orthogonal Multiple Access, NOMA)技术的应用能够更进一步减少多任务卸载延迟的问题。对于MEC与AoI的结合方面，部分学者已经做了少量研究。为了实时捕获新鲜的状态信息，Li等人[4]利用无界约束马尔可夫方法解决状态采样和卸载处理的问题。Liu等人[5] 提出了一个基于状态更新的Q学习算法可以有效地解决如何获取状态更新的情况。Song等人[6]设计了一个包含单个MEC服务器和单个移动设备的系统，并提出一种轻权重任务调度和计算卸载算法以解决年龄最小化的问题。然而，上述文献只考虑到任务调度和计算资源分配对AoI的影响，没有涉及如何使用有限的频谱资源进一步减少AoI的情况。

非正交多址接入被认为是一种有效提高频谱利用率的方法，随着研究的深入，NOMA与AoI的结合逐渐引起了广泛的关注[7]。NOMA的思想是在同一频谱资源中多个用户可以同时被服务。在不需要更多无线资源的情况下，NOMA传输能够使多个用户的AoI下降[8]。文献[9]对NOMA和传统正交多址接入(Orthogonal Multiple Access, OMA)环境下的平均AoI做了比较，这是NOMA应用于AoI的第1次尝试。文献[10,11]研究了AoI在NOMA和OMA网络中的性能表现，根据AoI的定义，数据的生成和传输的调度都起着很关键的作用。NOMA被认为是处理大规模物联网部署的一种很有前途的技术[12,13]。NOMA的思想是利用功率域，使多个用户在同一时间或者频带内得到服务，与OMA相比，NOMA可以通过提高频谱利用率来降低AoI[14]。Pan等人[15] 研究了基于NOMA的状态更新系统，经过分析发现，在高信噪比和中信噪比的情况下，NOMA能够实现更新鲜的信息更新。Gómez等人[16] 设计了一个在源节点和目的节点之间的队列传输模型。在传输过程中，为了降低总体的AoI，NOMA被用来进行节点间的功率分配。将NOMA引入到AoI中，虽然考虑了频谱资源的限制，但是却忽视了边缘计算在降低AoI方面的作用。

目前，已经有越来越多的文献在不同的场景下最小化AoI，然而很少有文献在NOMA-MEC联合系统中研究AoI问题。因此，本文综合考虑计算资源和频谱资源对AoI的作用，在此基础上，引入干扰的问题，通过设计一种联合优化卸载因子和卸载功率的策略，让所有设备的平均更新代价最小。考虑到环境动态变化这种更现实的场景，采用多代理深度确定性策略梯度(Multi-Agent Deep Deterministic Policy Gradient, MADDPG)算法用于分配卸载任务量和卸载功率。最后给出性能仿真结果与分析。

2　系统模型

2.1　网络模型

如图1所示，在这个系统中考虑一个多设备的MEC系统，它由移动设备的集合$\mathcal{N}$、1个装配有MEC服务器的接入点(Access Point, AP)和1个干扰者J组成。其中，$N=|\mathcal{N}|$是移动设备的数量。移动设备D可以监测物理过程的当前状态(例如利用摄像机记录十字路口的交通情况)，在这个过程中需要进行数据处理。假设这个系统可以分成多个时隙，　$t\in\mathcal{T}=\{0,1,\cdots,T-1\}$，每个时隙的长度为$\tau$。在每个时隙开始时，设备可以从环境中采集当前的数据。移动设备可以选择处理原始数据按照本地计算或者卸载给边缘服务器计算的方式进行处理。$\alpha_i(t)\in[0,1]$表示设备i的卸载因子，当$\alpha_i(t)=0$时，表示数据在设备i处完全进行本地计算；$\alpha_i(t)=1$，表示数据完全卸载给AP进行计算。所有设备的卸载决策可以表示为$\alpha(t)=[\alpha_1(t),\alpha_2(t),\cdots,\alpha_N(t)]$。在卸载过程中，所有设备的卸载功率分配决策为$P(t)=[p_1(t),p_2(t),\cdots,p_N(t)]$，其中$p_i(t)\in[0,P_{\max}]$表示设备$i$的卸载功率，$P_{\max}$是最大卸载功率。在每个时隙，设备使用计数器记录获得的信息年龄[10]。

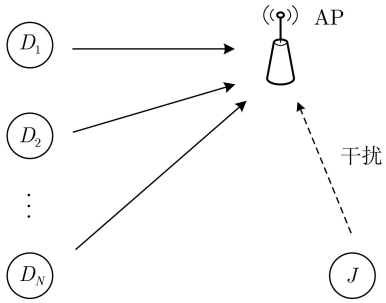

图 1 在多个设备中数据的安全传输

在卸载过程中，设备将受到干扰者的攻击。干扰者通过发射干扰信号阻碍设备与AP之间的通信，延长设备的卸载时间，从而使任务不能在一个时隙内完成，最终导致在一个时隙内状态更新失败。对于每个卸载设备，干扰者平均分配干扰功率。也就是说，当$\alpha_i(t)\neq 0$时，即部分数据或者全部数据通过卸载给AP进行处理时，在时隙t，干扰设备i的干扰功率为

$$p_J^i(t)=\frac{P_J}{n} \tag{1}$$

其中，P_J表示干扰者的总干扰功率，n表示选择卸载计算的设备数量。当$\alpha_i(t)=0$时，$p_J^i(t)=0$。即数据完全本地计算时，干扰者不发送干扰功率。

2.2 计算模型

在计算模型中，设备可以通过本地计算或者卸载计算来处理原始数据。

2.2.1 本地计算

假设每个设备装配1个本地处理器，它可以进行一些必要的计算，如果设备选择通过本地计算的方式处理数据，那么本地计算的时延$T_i^{\mathrm{loc}}(t)$可以表示为

$$T_i^{\mathrm{loc}}(t)=\frac{C_i}{f_i^{\mathrm{loc}}(t)} \tag{2}$$

其中，C_i表示在设备i计算这个原始数据总CPU周期数，$f_i^{\mathrm{loc}}(t)$是设备i在本地计算时CPU的频率。

本地计算时，设备i消耗的能耗$E_i^{\mathrm{loc}}(t)$为

$$E_i^{\mathrm{loc}}(t)=\mathrm{e}[f_i^{\mathrm{loc}}(t)]^2C_i \tag{3}$$

其中，e是能量系数，取决于芯片结构，设置$\mathrm{e}=10^{-27}$。

2.2.2 卸载计算

首先，采用NOMA的方式将多个数据卸载给边缘服务器。假设多个设备的信道增益满足

$$|h_1(t)|^2\geq|h_2(t)|^2\geq\cdots\geq|h_N(t)|^2 \tag{4}$$

其中，$|h_i(t)|^2$表示设备i与AP之间的信道增益。在上行NOMA中，具有高信道增益的设备先被解码，将较低信道增益的设备和干扰信号视为干扰。考虑设备i在时隙t选择卸载数据，它的传输速率$R_i(t)$可以表示为

$$R_i(t)=B\log_2\left(1+\frac{p_i(t)|h_i(t)|^2}{\sum\limits_{m=i+1}^{N}p_m(t)|h_m(t)|^2+L+\sigma^2}\right) \tag{5}$$

其中，B是系统的带宽，$p_i(t)$表示设备i的传输功率，$L=p_j^i(t)\left|h_j^i(t)\right|^2$代表干扰者对设备$i$的干扰，$\sigma^2$表示加性噪声功率，$h_m(t)$表示其他信道的信道增益。

在数据卸载过程中，卸载时间$T_i^{\mathrm{off}}(t)$可以表示为

$$T_i^{\mathrm{off}}(t)=\frac{\alpha_i(t)D_i}{R_i(t)} \tag{6}$$

其中，D_i是输入的总数据量。在时隙t，设备i的卸载能耗$E_i^{\mathrm{off}}(t)$为

$$E_i^{\mathrm{off}}(t)=p_i(t)T_i^{\mathrm{off}}(t) \tag{7}$$

然后，将多个设备的数据传输到边缘服务器后，边缘服务器将进行计算，定义MEC服务器可利用的总计算资源为$F(t)$，$f_i^{\mathrm{ex}}(t)$代表边缘服务器分配给设备i的计算资源。因此，在边缘服务器进行数据处理的时间$T_i^{\mathrm{ex}}(t)$为

$$T_i^{\mathrm{ex}}(t)=\frac{C_i}{f_i^{\mathrm{ex}}(t)} \tag{8}$$

其中，C_i表示处理设备i的数据时所需的总CPU的周期数。

最后，经过边缘服务器处理后，可以得到计算结果。由于计算结果的数据量很小，传输速度较快，因此，传输时延可以忽略不计。

因此，设备i处理任务的时延可以表示为

$$T_i(t)=\begin{cases}T_i^{\mathrm{loc}}(t), & \alpha=0\\ \max\{T_i^{\mathrm{loc}}(t),T_i^{\mathrm{off}}(t)+T_i^{\mathrm{ex}}(t)\}, & 0<\alpha<1\\ T_i^{\mathrm{off}}(t)+T_i^{\mathrm{ex}}(t), & \alpha=1\end{cases} \tag{9}$$

设备i的能耗表示为

$$E_i(t)=(1-\alpha_i(t))E_i^{\mathrm{loc}}(t)+\alpha_i(t)E_i^{\mathrm{off}}(t) \tag{10}$$

2.3 状态更新模型

在每个时隙t，设备通过处理计算任务来获得状态更新。如果计算任务能在这个时隙内完成，则状态信息被更新；否则，设备没有状态更新。在这部分，利用信息年龄来测量状态更新的新鲜度。在这个多设备MEC系统中，AoI反映在设备处生成最新被执行的任务，到被处理，最终在设备处获得计算结果所经过的时间。采用$A_i(t)=t-\theta_i(t)$表示设备i的信息年龄。其中$\theta_i(t)$是指设备i产生最新任务的时间戳。信息年龄的演变展示在图2中。

在每个时隙开始的时刻，设备i从周围的环境中采样一个需要处理的计算任务，T_t表示时隙t的起始时刻，也是采样计算任务的时刻。计算任务通过3种形式进行处理：本地计算、部分卸载、完全卸载。在时隙t，任务的时延用式(9)表示。因此，在时隙t获得计算结果的时刻可以用$D_t=T_t+T_i(t)$表示。在这个系统中考虑计算任务需要在一个时隙内完成，也就是说，$T_i(t)\leq\tau$。当设备i 接收到计算结果后，在下一个采样时刻T_{t+1}开始前，设备i处可能会产生一个等待时间$Z_t\in[0,\tau]$，所以下一

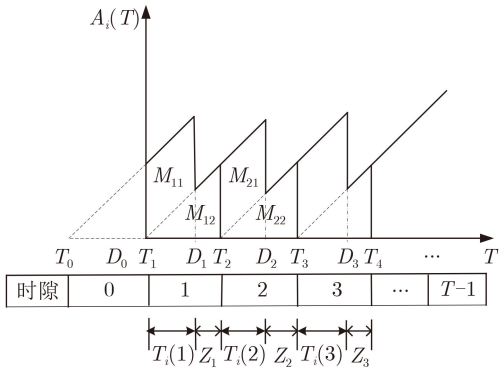

图 2 信息年龄的演变

个采样时刻可以表示为$T_{t+1}=D_t+Z_t$。在图2中，可以看出在等待下一次采样和处理数据时，AoI是不断增加的。当计算结果被传递给设备时，AoI开始急速向下跳变。特别地，在时隙t，时刻D_t的AoI可以表示为$A_i(D_t)=D_t-T_t=T_i(t)$。

基于平均信息年龄和干扰代价，设备i在时隙t的更新代价$u_i(t)$为

$$u_i(t)=\bar{A}_i(t)+wp_j^i(t) \tag{11}$$

其中，$\bar{A}_i(t)=\dfrac{1}{\tau}\displaystyle\int_{T_t}^{T_{t+1}}A_i(t)\mathrm{d}t$是设备$i$在时隙$t$的平均信息年龄。$T_t$表示时隙$t$的起始时刻，$T_{t+1}$代表时隙$t+1$的起始时刻，也是时隙$t$的结束时刻。$w$是抵抗单位干扰功率的代价。为了使每个设备的更新代价最小，本文通过最小化平均更新代价来实现。每个设备的平均更新代价表示为

$$\begin{aligned}\bar{U}&=\lim_{T\to\infty}\frac{1}{NT}E\left\{\sum_{t=0}^{T-1}\sum_{i=1}^{N}u_i(t)\right\}\\&=\frac{1}{N}\sum_{i=1}^{N}\left[\bar{A}_i+\frac{1}{T}\sum_{t=0}^{T-1}wp_j^i(t)\right]\end{aligned} \tag{12}$$

其中，$\bar{A}_i=\lim\limits_{T\to\infty}\dfrac{1}{T}\displaystyle\int_0^T A_i(t)\mathrm{d}t$表示设备$i$的长期平均信息年龄，与图2围成的面积有关。为了计算$\bar{A}_i$，将平均AoI分解成不同的区域，如图2所示。其中M_{t1}围成的平行四边形面积，可以表示为

$$M_{t1}=\left[T_i(t-1)+Z_{t-1}\right]T_i(t)=\tau T_i(t) \tag{13}$$

M_{t2}围成的是三角形区域，表示为

$$M_{t2}=\frac{1}{2}\left[T_i(t)+Z_t\right]^2=\frac{1}{2}\tau^2 \tag{14}$$

因此，平均AoI可以用下面的公式进行计算

$$\begin{aligned}\bar{A}_i=&\frac{\sum\limits_{t\to\infty}\left(M_{t1}+M_{t2}\right)}{\sum\limits_{t\to\infty}\left(T_i(t)+Z_t\right)}\\=&\sum_{t\to\infty}\left(T_i(t)+\frac{1}{2}\tau\right)\end{aligned} \tag{15}$$

3　优化问题

在本部分，优化的目标是在计算资源、处理时延和用户能耗的约束下最小化平均更新代价，即

$$\left.\begin{aligned}&\min_{\alpha,P}\ \bar{U}\\&\text{s.t.C1}:\alpha_i(t)\in[0,1],\ \forall i\in\mathcal{N},t\in\mathcal{T}\\&\quad\text{C2}:p_i(t)\in(0,P_{\max}],\ \forall i\in\mathcal{N},t\in\mathcal{T}\\&\quad\text{C3}:T_i^{\text{off}}(t)+T_i^{\text{ex}}(t)\le\tau,\ \forall i\in\mathcal{N},t\in\mathcal{T}\\&\quad\text{C4}:T_i^{\text{loc}}(t)\le\tau,\ \forall i\in\mathcal{N},t\in\mathcal{T}\\&\quad\text{C5}:T_i^{\text{off}}(t)<\tau,\ \forall i\in\mathcal{N},t\in\mathcal{T}\\&\quad\text{C6}:\sum_{i=1}^{N}E_i(t)\le E_{\max},\ \forall t\in\mathcal{T}\\&\quad\text{C7}:\sum_{i=1}^{N}I[\alpha_i(t)\ne0]f_i^{\text{ex}}(t)\le F(t),\ \forall t\in\mathcal{T}\end{aligned}\right\} \tag{16}$$

由式(11)和式(14)可知，平均更新代价与处理时延有关，而处理时延受到卸载任务量和卸载功率的影响。因此，在多设备的MEC系统中，需要通过优化卸载决策和卸载功率来最小化平均更新代价。在上面的公式中，C1和C2分别表示卸载决策和卸载功率的取值范围。C3和C4分别表示利用卸载计算或者本地计算处理的任务需要在一个时隙内完成。C5表示通过优化变量来抵抗干扰攻击，降低传输时间，使卸载时间不超过一个时隙，确保在一个时隙内完成状态更新。C6保证所有设备的总能耗不超过设置的最大能耗。而C7保证分配给卸载设备的计算资源总和不超过MEC服务器的计算容量。由于在不同的时隙下，信道条件等变量是随着时间动态变化的，传统的优化方法难以解决动态变化的场景。而强化学习能有效地解决这一问题。因此，采用强化学习算法来优化卸载决策和卸载功率，从而使目标函数最小。

4　MADDPG算法

强化学习是单个代理与未知环境相互交互，使长期奖励最大化的一种有效方法。通过不断地尝试，它可以让单个代理学习到最优的行为。强化学习由3个必要的变量组成：状态，动作，奖励。在每次迭代过程中，代理将从环境中选择当前的状态信息，将它作为输入值，然后选择一个动作，环境会根据选择的动作值反馈给代理一个奖励，用来评价当前动作的好坏。通过反复的试错，代理会倾向选择使长期奖励增加的动作[17]。

在多设备的MEC系统中，本文将每个设备视为一个代理，设备之外的一切被视为环境。考虑到卸载速率、设备的总能耗和MEC服务器计算容量的影响，其他设备的决策会对当前代理产生影响。

由此可以看出，欲最小化平均更新代价，需要多个代理的相互协作才能实现。然而，在多代理的环境中，传统的强化学习是不适用的。这是因为在传统强化学习中，每个代理只考虑最大化自身的奖励，没有考虑其他代理的影响。针对这一问题，多代理强化学习可以提供一个有效的解决方法。多代理强化学习允许多个代理通过相互协作来实现它们的目标。结合当前场景，状态、动作和奖励对应如下：

状态：在时隙t，代理i观察网络的情况，并且选择下面的参数构成网络的状态。

$\bar{A}_i(t)$：代理i在时隙t的平均信息年龄。

$\varphi(t)$：MEC服务器剩余的计算容量。其中$\varphi(t)=F(t)-\sum_{i=1}^{N} I[\alpha_i(t)\neq 0]f_i^{\mathrm{ex}}$。

$E(t)$：N个用户剩余的能耗。$E(t)$用公式表示为$E(t)=E_{\max}-\sum_{i=1}^{N}\left[(1-\alpha_i(t))E_i^{\mathrm{loc}}(t)+\alpha_i(t)E_i^{\mathrm{off}}(t)\right]$。

因此，代理i的状态可以表示为$s_i(t)=(A_i(t),\varphi(t),E(t))$，让$S$表示状态空间，它由所有代理的状态组成，$S=\{s_1(t),s_2(t),\cdots,s_N(t)\}$。

动作：在时隙t，代理i的动作用$a_i(t)$表示，它由下面几部分组成：

$\alpha_i(t)$：代理i在时隙t的卸载因子。$\alpha_i(t)\in[0,1]$，如果$\alpha_i(t)=0$表示代理i选择完全本地计算；$\alpha_i(t)=1$是代理i选择完全卸载给AP进行计算；如果$0<\alpha_i(t)<1$代表代理i选择部分卸载。

$p_i(t)$：代理i在时隙t的卸载功率。它的取值范围是$p_i(t)\in(0,P_{\max}]$。

因此，代理i的动作用公式表示为$a_i(t)=(\alpha_i(t),p_i(t))$，利用$A$代表动作空间，它由所有代理的动作组成。$A=\{a_1(t),a_2(t),\cdots,a_N(t)\}$。

奖励：在选择完动作$a_i(t)$后，代理i将会获得瞬时奖励值$r_i(t)$。一般来说，奖励值应该和目标函数(式(11))是相关的。在这个系统中优化问题是最小化平均更新代价，而多代理强化学习的目标是获得最大奖励值。因此，奖励与目标函数是负相关的，设置代理i在时隙t的奖励为

$$r_i(t)=-u_i(t) \tag{17}$$

其中，$u_i(t)$是代理i在时隙t的代价。

在本文中，由于动作的取值是连续的，需要采用基于策略的算法进行求解。考虑到有大量的设备需要处理自身的计算任务，因此，代理的数量是非常大的。基于策略的演员-评论家算法(Actor-Critic, AC)在单代理的环境中表现良好，但是随着代理数量的增加，方差也会变大，所以不适用于多代理的环境。而多代理深度确定性策略梯度算法是AC算法的一种变体，通过让智能体之间集中训练以及分布运行，它可以处理动态环境中环境与代理相互交互的问题，在代理i做决策时，会考虑其他代理的影响。通过多个代理间协作，共同最大化奖励值。针对上述优势，采用MADDPG算法来寻找最优的动作值，从而达到最小化目标函数的目的。

在MADDPG算法中，利用经验回放机制降低样本之间的相关性。通过代理与环境的交互，可以获得经验序列(s_t,a_t,r_t,s_{t+1})，其中s_t,a_t,r_t分别对应状态、动作和奖励。s_{t+1}表示下一个状态。所有代理的经验被存储在经验回放内存D中。在训练过程中，从D中随机抽取小批经验序列进行学习。MADDPG算法主要是由AC的框架组成。在演员A中，它主要由在线策略网络和目标策略网络组成。确定性策略μ直接从每步的动作中获得。在评论家C中，它也主要由两个网络组成：在线Q网络和目标Q网络。对于演员框架，在线策略网络的更新主要由策略梯度来完成，策略梯度的表达式为

$$\nabla_{\theta^\mu}J(\theta^\mu)\approx\frac{1}{M}\sum_i\nabla_a Q(s_i,a|\theta^Q)|_{a=\mu(s_i)}\nabla_{\theta^\mu}\mu(s_i|\theta^\mu) \tag{18}$$

其中，$Q(s_i,a|\theta^Q)$为第i个智能体的动作值函数，M为所有策略数目的总和。

在评论家框架中，在线Q网络的参数由损失函数进行更新，损失函数的表达式为

$$L=\frac{1}{M}\sum_i\left(y_i-Q(s_i,a_i|\theta^Q)\right)^2 \tag{19}$$

其中，$y_i=r_i+\gamma Q'(s_{i+1},\mu'(s_{i+1}|\theta^{\mu'})|\theta^{Q'})$，$\gamma$表示折扣因子。

对于演员-目标策略网络和评论家-目标Q网络分别用式(19)和式(20)表示

$$\theta^{\mu'}\leftarrow\tau\theta^\mu+(1-\tau)\theta^{\mu'} \tag{20}$$

$$\theta^{Q'}\leftarrow\tau\theta^Q+(1-\tau)\theta^{Q'} \tag{21}$$

其中，τ代表更新参数，满足$\tau\ll 1$，用来改善训练的稳定性。用于多设备NOMA-MEC系统的MADDPG算法训练如下：

(1) 初始化演员网络$\mu(s|\theta^\mu)$，评论家网络$Q(s,a|\theta^Q)$；

(2) 初始化权重θ^μ, θ^Q, $\theta^{\mu'}\leftarrow\theta^\mu$和$\theta^{Q'}\leftarrow\theta^Q$；

(3) 初始化随机过程M；

(4) 初始化状态s_0；

(5) for (每个时刻$t=0,1,\cdots,T$) do；

(6) 基于当前策略选择动作$a_t=\mu(s_t|\theta^\mu)+M_t$；

(7) 执行动作a_t，得到奖励r_t，并转移到下一个状态s_{t+1}；

(8) 存储经验序列(s_t, a_t, r_t, s_{t+1})到内存D；

(9) 从D中抽取小批量的样本进行训练；

(10) 根据式(17)更新演员-在线策略网络参数；

(11) 设定$y_i = r_i + \gamma Q'(s_{i+1}, \mu'(s_{i+1}|\theta^{\mu'})|\theta^{Q'})$；

(12) 根据式(18)更新评论家-在线Q网络的参数；

(13) 分别用式(19)和式(20)更新演员-目标策略网络和评论家-目标Q网络参数；

(14) end for。

5　仿真结果分析

在该部分，本文考虑不同工作模式、卸载功率和不同算法对平均更新代价的影响。在这个场景中，设定设备被随机地分布在200 × 200 m^2的区域内，与服务器相连的AP位于该区域的中心，干扰者在AP的附近。输入任务的大小Di(kbit)服从(100, 500)之间的均匀分布，处理1 bit数据所需的CPU周期数为2×10^3 cycle/bit。信道带宽为2 MHz，相应的噪声功率$\sigma^2 = 3 \times 10^{-13}$。另外，可利用的MEC服务器的计算容量$F(t)$设置为10 GHz。在本地计算阶段，每个设备的CPU频率为0.2 GHz。在传输过程中，单位干扰功率的代价w为0.1，总干扰功率P_J设置为20 W。仿真参数如表1所示。

图3展示了在设备数量设置为10，3种卸载因子的作用下，不同MEC计算容量对平均更新代价的影响。这3种卸载因子分别表示本地计算($\alpha = 0$)，部分卸载($\alpha = 0.5$)和完全卸载($\alpha = 1$)。由图3可以看出，随着MEC服务器计算容量的增加，部分卸载和完全卸载的长期平均代价都逐渐减小，而本地计算的长期平均代价保持不变。这是因为当MEC服务器的计算容量增加时，更多的设备可以通过将计算任务卸载给MEC服务器处理来获得状态更新。并且，对于仅本地计算来说，每个设备的状态更新不受MEC服务器计算容量的影响。因此，通过部分卸载的方式和适当地增加MEC服务器的计算容量，可以有效地降低平均更新代价。

然后，考虑在部分卸载(卸载因子为0.5)的情况下，利用3种不同的方案去优化卸载功率从而使平均更新代价最小。这3种方案表示如下：

(1) MADDPG算法，即主要应用的优化方案。

(2) 演员-评论家算法(AC算法)：每个设备不知道其他设备的信息，在训练过程中，只知道自身的本地信息。

(3) Q学习算法：每个设备不知道其他设备的信息，适用于小规模离散动作空间的优化。

图4展示了在固定用户数量下，迭代次数和平均更新代价的关系。从图4可以看出，随着迭代次数的增加，平均更新代价逐渐减小。除此之外，MADDPG算法在降低平均更新代价方面优于其他两种方案。这是因为MADDPG算法考虑到多个代理之间的相互协作，通过代理间的共同作用，最大化奖励值。而AC算法和Q学习算法没有考虑到设备间的相互影响，只考虑自身的状态信息。从图4还可以看出，MADDPG算法的平均更新代价分别比AC算法和Q学习算法降低了37.5%和53.1%。

图5表示不同设备数量对平均更新代价的影响。当设备数量在10～100逐渐增加时，3种算法的平均更新代价也是逐渐增加的。这是因为MEC服务器计算容量有限，随着设备数量的增加，每个设备获得的计算资源减少，因此导致处理时间增加，进而使平均更新代价增大。通过对图中数据分析可以发现，适当地减少用户数量，有利于降低平均更新代价。

表 1　仿真参数设置

CPU周期	信道带宽	计算容量	CPU频率	干扰功率的代价	总干扰功率
2×10^3cycle/bit	2 MHz	10 GHz	0.2 GHz	0.1	20 W

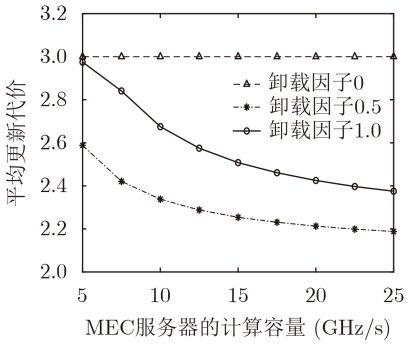

图 3 计算容量对平均更新代价的影响

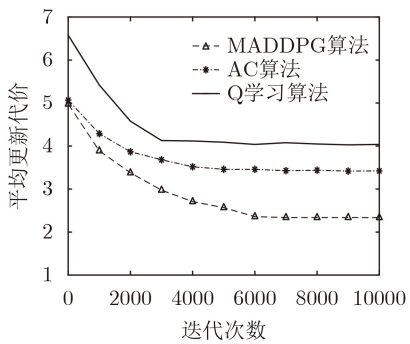

图 4 迭代次数和平均更新代价的关系

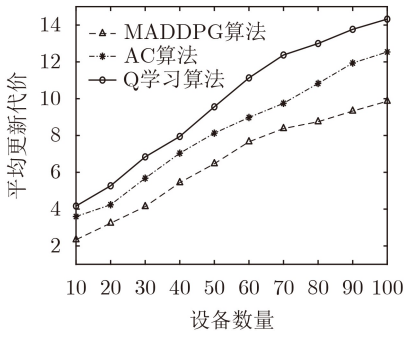

图 5 不同设备数量对平均更新代价的影响

6 结束语

本文基于NOMA-MEC联合系统，考虑到信息新鲜度的影响，对多设备单边缘计算服务器场景进行了研究。基于MADDPG算法，建立了最小化平均信息年龄的优化问题，提出一种寻优的卸载因子和卸载功率策略。仿真结果表明，利用部分卸载的方式，在降低平均更新代价方面效果最好。同时，与其他方案相比，采用MADDPG算法和降低设备数量均可有效地降低平均更新代价。提出的寻优的卸载因子和卸载功率策略可以很好地降低设备的信息更新代价，大大提高了设备的更新效率。

参 考 文 献

[1] KAUL S, YATES R, and GRUTESER M. Real-time status: How often should one update?[C]. 2012 IEEE INFOCOM, Orlando, USA, 2012: 2731–2735. doi: 10.1109/INFCOM.2012.6195689.

[2] ZOU Peng, OZEL O, and SUBRAMANIAM S. Trading off computation with transmission in status update systems[C]. The IEEE 30th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), Istanbul, Turkey, 2019: 1–6. doi: 10.1109/PIMRC.2019.8904244.

[3] MAO Yuyi, YOU Changsheng, ZHANG Jun, *et al*. A survey on mobile edge computing: The communication perspective[J]. *IEEE Communications Surveys & Tutorials*, 2017, 19(4): 2322–2358. doi: 10.1109/COMST.2017.2745201.

[4] LI Rui, MA Qian, GONG Jie, *et al*. Age of processing: Age-driven status sampling and processing offloading for edge-computing-enabled real-time IoT applications[J]. *IEEE Internet of Things Journal*, 2021, 8(19): 14471–14484. doi: 10.1109/JIOT.2021.3064055.

[5] LIU Long, QIN Xiaoqi, TAO Yunzheng, *et al*. Timely updates in MEC-assisted status update systems: Joint task generation and computation offloading scheme[J]. *China Communications*, 2020, 17(8): 168–186. doi: 10.23919/JCC.2020.08.014.

[6] SONG Xianxin, QIN Xiaoqi, TAO Yunzheng, *et al*. Age based task scheduling and computation offloading in mobile-edge computing systems[C]. 2019 IEEE Wireless Communications and Networking Conference Workshop (WCNCW), Marrakech, Morocco, 2019: 1–6. doi: 10.1109/WCNCW.2019.8902529.

[7] ZENG Jie, LV Tiejun, LIU Renping, *et al*. Investigation on evolving single-carrier NOMA into multi-carrier NOMA in 5G[J]. *IEEE Access*, 2018, 6: 48268–48288. doi: 10.1109/ACCESS.2018.2868093.

[8] MAATOUK A, ASSAAD M, and EPHREMIDES A. Minimizing the age of information: NOMA or OMA?[C]. 2019 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), Paris, France, 2019: 102–108. doi: 10.1109/INFCOMW.2019.8845254.

[9] WANG Qian, CHEN He, LI Yonghui, *et al*. Minimizing age of information via hybrid NOMA/OMA[C]. 2020 IEEE International Symposium on Information Theory (ISIT), Los Angeles, USA, 2020: 1753–1758. doi: 10.1109/ISIT44484.2020.9174163.

[10] CHEN Xiang, GONG Fengkui, LI Guo, *et al*. User pairing and pair scheduling in massive MIMO-NOMA systems[J]. *IEEE Communications Letters*, 2018, 22(4): 788–791. doi: 10.1109/LCOMM.2017.2776206.

[11] DING Zhiguo, LIU Yuanwei, CHOI J, *et al*. Application of non-orthogonal multiple access in LTE and 5G networks[J]. *IEEE Communications Magazine*, 2017, 55(2): 185–191. doi: 10.1109/MCOM.2017.1500657CM.

[12] SAITO Y, KISHIYAMA Y, BENJEBBOUR A, *et al*. Non-Orthogonal Multiple Access (NOMA) for cellular future radio access[C]. The IEEE 77th Vehicular Technology Conference (VTC Spring), Dresden, Germany, 2013: 1–5. doi: 10.1109/VTCSpring.2013.6692652.

[13] ZHOU Bo and SAAD W. Joint status sampling and updating for minimizing age of information in the internet of things[J]. *IEEE Transactions on Communications*, 2019, 67(11): 7468–7482. doi: 10.1109/TCOMM.2019.2931538.

[14] YU Yuehua, CHEN He, LI Yonghui, *et al*. On the

performance of non-orthogonal multiple access in short-packet communications[J]. *IEEE Communications Letters*, 2018, 22(3): 590–593. doi: 10.1109/LCOMM.2017.2786252.

[15] PAN Haoyuan, LIANG Jiaxin, LIEW S C, *et al*. Timely information update with nonorthogonal multiple access[J]. *IEEE Transactions on Industrial Informatics*, 2021, 17(6): 4096–4106. doi: 10.1109/TII.2020.3006061.

[16] GÓMEZ J T, MORALES-CÉSPEDES M, ARMADA A G, *et al*. Minimizing age of information on NOMA communication schemes for vehicular communication applications[C]. The 12th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP), Porto, Portugal, 2020: 1–6. doi: 10.1109/CSNDSP49049.2020.9249492.

[17] ELGABLI A, KHAN H, KROUKA M, *et al*. Reinforcement learning based scheduling algorithm for optimizing age of information in Ultra Reliable low latency networks[C]. 2019 IEEE Symposium on Computers and Communications (ISCC), Barcelona, Spain, 2019: 1–6. doi: 10.1109/ISCC47284.2019.8969641.

李保罡：男，副教授，研究方向为无线通信、工业互联网、能源互联网、大数据分析等.

石　泰：男，硕士生，研究方向为边缘计算、物理层安全等.

陈　静：女，助理工程师，研究方向为互联网部运营监测.

李诗璐：女，硕士，研究方向为边缘计算.

王　宇：男，硕士生，研究方向为物理层安全.

张天魁：男，教授，研究方向为移动边缘计算、无线资源管理、未来网络融合与管理等.

责任编辑：余　蓉