

边缘计算环境下任务卸载与资源调度优化策略

赵智宇

北京中芯标准科技有限公司天津分公司 天津市 300010

【摘要】：边缘计算就是将网络边缘的计算、存储资源用在终端设备上，给资源受限的终端设备提供低时延、高带宽的服务。任务卸载决策和资源调度的联合优化属于决定系统性能的重要部分，在动态环境适应性、多服务器协同以及能耗和时延的均衡方面还存在着改进的空间。本文提出了一种基于深度强化学习的任务卸载和资源调度联合优化的方法，把问题建模成马尔可夫决策过程，用包含任务特征、信道状态、服务器负载信息的系统状态来描述，设计出以加权时延、能耗最小化为目标的奖励函数，用改进的双延迟深度确定性策略梯度算法进行策略学习。从仿真实验结果可知，在多用户、多服务器的场景中，本文所提的方法在系统总成本方面比基线方法有所降低，在动态信道环境下的收敛性和稳定性较好。

【关键词】：边缘计算；任务卸载；资源调度；深度强化学习；联合优化

0 引言

伴随着物联网设备的大量部署以及第五代移动通信技术的普遍应用，移动终端上所承载的应用类型也越来越繁杂，增强现实、自动驾驶、工业自动控制等场景对计算能力以及响应速度有着很高的要求。边缘计算把计算能力下沉到靠近数据源头的网络边缘，让终端设备可以将计算任务卸载到就近的边缘服务器上执行，从而减轻核心网络的传输压力和降低服务时延。

任务卸载和资源调度属于边缘计算研究中的两个关键问题，两者之间存在很强的联系。卸载决策包含是否卸载以及卸载到哪个服务器的选择，资源调度就是决定各个边缘服务器怎样分配有限的计算资源给不同的任务。由于任务到达的随机性、信道状态的时变性、服务器负载的动态性，这两个问题互相影响，必须一起考虑。在这样的复杂环境下，怎样才能达到高效的、自适应的联合优化呢？目前仍然是学术界和工业界共同关注的难点。

现有的研究工作大致可以分为三条路径。第一类用确定性优化的方法，把问题抽象成混合整数非线性规划模型，用分支定界或者动态规划得到最优解，但是这类方法的计算复杂度随着问题规模呈指数增长，不能满足实时决策的要求。第二类用启发式或者元启发式算法，粒子群优化、遗传算法等，在求解速度上有所提高，但是对环境变化的反应比较慢，没有在线学习的能力。第三类是利用深度强化学习，用智能体和环境的不断交互来逼近最优策略，在动态卸载问题上已经取得了较好的效果。

本文在已有成果的基础上，提出一种基于改进双延迟深度确定性策略梯度（TD3）的任务卸载和资源调度联合优化策略。主要工作有三部分，分别是建立融合多维动态信息的系统状态表征，使智能体可以更好地感知环境的变化，设计时延和能耗兼顾的加权奖励函数，实现多目标间的折中，用仿真实验验证

所提策略在动态多服务器环境中的有效性、稳定性。

1 系统模型与问题描述

1.1 系统架构

在由多个移动终端设备、若干个边缘服务器和远端云数据中心组成的三层计算卸载系统中，每一个时隙都会有终端产生一个计算任务，任务可以用数据量大小、所需计算资源量和最大可接受时延这三个属性来描述。终端可以自己执行这个任务，也可以把任务通过无线链路传送到一个边缘服务器上执行，如果边缘服务器负载过大，还可以将任务传送到云端进行处理。

每个边缘服务器都有一些计算资源，这些资源要分配给多个卸载任务。终端和边缘服务器之间无线信道的状态会随时间而变化，上行传输速率受信道增益、发射功率和噪声水平等影响。系统架构如图1所示，该图直观地展示了从终端设备到边缘节点再到云端的任务流转路径。

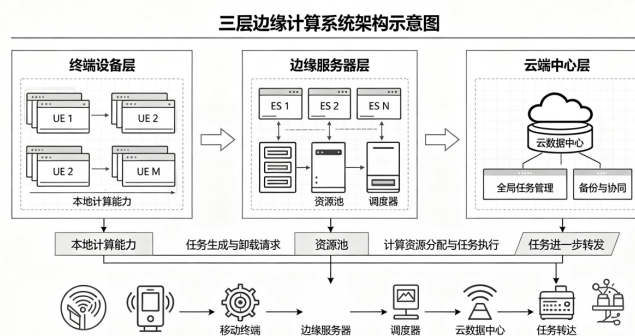


图1 三层边缘计算系统架构示意图

1.2 任务卸载与资源分配过程

在每一个时隙中，如果某个终端选择将任务卸载到某个边缘服务器上，那么任务处理时延就由两部分组成，即把任务数据上传到该服务器的传输时延，也就是任务数据量除以上行传输速率，任务在该服务器上计算的时延，也就是任务所需计算

周期数除以服务器分配给该任务的计算资源量。如果任务在本地执行,则只产生本地计算时延,即所需计算周期数除以终端本地处理能力。

系统能耗包括传输能耗和计算能耗。传输能耗为终端发射功率乘以传输时延,边缘服务器的计算能耗和它所分配到的计算资源量成平方关系,能耗系数与服务器硬件特性有关。每一个任务的完成时延要符合它的最大容忍约束,否则就属于服务质量违规。

1.3 优化目标与约束

本文的优化目标就是在满足各个任务的时延约束以及边缘服务器计算容量约束的情况下,使系统在长时间运行中平均加权总成本最小,总成本是所有任务时延和能耗的加权和。权重系数是根据具体的应用场景来确定时延、能耗所占的比重。该问题本质上就是混合整数非线性规划问题,决策变量既有离散的卸载选择又有连续的资源分配比例,所以很难在多项式时间内得到全局最优解,需要设计高效启发式或者学习型的方法。

2 基于深度强化学习的联合优化方法

2.1 马尔可夫决策过程建模

为了利用深度强化学习来解决上述优化问题,首先要将其转化为马尔可夫决策过程,马尔可夫决策过程包含状态、动作、奖励这三个主要部分。

状态空间的设计考虑了影响决策的主要因素,即终端当前任务数据量、计算周期、最大容忍时延、终端与边缘服务器之间上行传输速率、边缘服务器剩余可用计算资源等。这些状态信息一起成为智能体作出卸载和调度决策的依据。

动作空间分为两大部分,离散的卸载决策,也就是每一个任务是否需要卸载,以及卸载到哪一台服务器上;连续的资源分配方案,即每个服务器为它所承担的任务分别分配多少计算资源。由于混合动作空间的处理比较困难,因此本文采用解耦的方式,即先由神经网络输出卸载偏好,然后根据卸载结果生成连续的资源分配值。

奖励函数为负的加权总成本,对违反时延约束的进行惩罚。该种方式使智能体在最小化时延和能耗的同时尽量满足服务质量的要求。

2.2 改进的TD3算法

确定性策略梯度法对于连续动作问题来说容易出现Q值过高估计的情况,进而影响到策略的质量。本文使用双延迟深度确定性策略梯度(TD3)算法对Q值进行优化,有三个关键技术可以提高训练的稳定性,第一,用两个Critic网络分别计算Q值,取较小的一个作为目标值,抑制过估计偏差;第二,延迟Actor网络的更新频率,使Critic网络在Actor更新之

前有更充分的训练;第三,给目标动作加上裁剪后的噪声,实现策略平滑正则化,防止对确定性策略的过度依赖。

训练过程使用离线学习的方式,智能体同环境交互产生的经验元组(包含当前状态、动作、奖励、下一状态等信息)被存入经验回放池,训练时从池中随机抽取少量样本来更新网络参数。为了兼顾探索和利用,在动作选择的时候加入高斯噪声,随着训练的进行逐渐减小噪声强度,使策略由探索为主转变为利用为主。算法在每一个回合里,会先做状态观测、动作选择、环境反馈、经验存储以及网络更新这些工作,直到奖励曲线达到收敛为止。

3 实验设计与结果讨论

3.1 仿真环境与参数配置

为了检验所提出的方法是否有效,建立了一个仿真实验平台。实验条件是终端数量为20个,边缘服务器数量为5个,各个服务器的计算能力都在10到20吉赫兹之间均匀分布。任务数据量在0.5到5兆字节之间随机产生,所需要的计算时间在0.5到2吉周期之间,最大的容忍时延为0.5到3秒。无线信道用瑞利衰落模型来模拟,信噪比在5到25分贝之间随机变化。时延权重、能耗权重分别为0.6、0.4,时延优先、能耗优先并重。

对比方法有全部任务在终端本地执行、随机选择卸载目标、标准深度确定性策略梯度算法、粒子群与灰狼混合优化算法。性能评价指标有平均任务时延、平均系统能耗、加权总成本和时延约束满足率。

3.2 收敛性能分析

从训练过程中奖励曲线变化趋势可知,提出TD3算法前200个回合内平均奖励增长较快,说明智能体可以很快地通过和环境的交互来掌握有效的卸载、调度策略。200到500回合之间奖励曲线有小幅度的波动,是正常的探索和利用权衡的表现。经过大约500回合的训练之后,奖励值趋于稳定,波动幅度小于5%,说明算法有较好的收敛性。与标准DDPG相比,TD3的收敛速度慢但是收敛值低,是因为它采用了更保守的更新策略来减小训练过程中出现的方差。

3.3 不同方法性能对比

表1给出了各个对比方法在500次独立重复实验中得到的平均性能。从数据可知,本地执行方案由于缺少卸载协同,时延、能耗均处于较高水平,时延约束满足率为62.3%。随机卸载方案虽然有所改善,但是决策的盲目性导致资源利用不充分。PSO-GWO混合优化方法在离线优化场景下表现还可以,但是它对于环境变化的响应存在延迟。标准DDPG算法在线学习得到较好的性能,本文提出的TD3策略在平均时延、平均能耗、加

权总成本上都比所有的对比方法好。就加权总成本而言，相比于 DDPG 来说，TD3 的加权总成本降低了大约 18.4%；就时延约束满足率来说，TD3 的时延约束满足率比 DDPG 高很多，达到了 94.1%。因此，它在满足服务质量方面有明显的优势。

表1 不同方法的系统性能对比（平均值）

方法	平均时延（秒）	平均能耗（焦耳）	加权总成本	时延约束满足率
本地执行	2.84	18.7	9.18	62.3%
随机卸载	1.96	15.2	7.26	71.5%
PSO-GWO	1.42	12.8	5.97	84.2%
标准DDPG	1.23	11.5	5.34	89.7%
本文TD3策略	1.07	9.8	4.36	94.1%

3.4 动态环境下的适应性评估

为了检验策略在非平稳环境中的鲁棒性，在测试阶段人为引入了两种突发扰动：一是任务到达率瞬时增加两倍，二是无线信道质量突然恶化。实验结果表明，本文 TD3 策略在扰动发生后的 3 到 5 个时隙内就可以将系统成本恢复到接近扰动前的水平，而标准 DDPG 和 PSO-GWO 分别需要 8 到 12 个和 10 到 15 个时隙。主要是由于 TD3 算法的双网络结构以及策略平滑机制，使得价值估计更加准确，策略更新更加平稳，在环境突然变化的时候有较好的适应性。从整体上看，所提出的方法对于动态的边缘环境来说是稳定、快速的。

4 结论

本文针对边缘计算中任务卸载和资源调度的联合优化问题，提出了一个基于改进 TD3 深度强化学习的方法。将问题建

模成马尔可夫决策过程，用状态空间和加权奖励函数来体现系统的动态性，从而达到卸载决策和资源分配的协同优化。仿真实验在多用户多服务器环境下验证了该方法的有效性，结果表明，它在系统加权总成本、时延约束满足率、环境变化适应性等各方面都比本地执行、随机卸载、混合元启发式算法、标准深度确定性策略梯度等方法要好。

尽管本文的方法取得了一些积极的结果，但是仍然存在很多值得进一步研究的地方。目前模型假定所有的任务都是可观察的，并且决策是集成的，将来可以考虑多边缘服务器之间的分布式协同决策，也可以使用联邦学习框架来保护用户的隐私。除此之外，把计算任务和通信资源联合感知，扩展到超可靠低时延通信场景中去，也是今后研究的一个方向。

参考文献

[1] 周绪, 苗辉, 杨静, 等. 边缘计算资源调度:历史、架构、建模与方法分析[J]. 计算机集成制造系统, 2025, 31(8): 2695-2726.

[2] 曾耀平, 李航淼, 陈世森. 基于DRL的多服务器接入边缘计算卸载与资源调度[J]. 计算机工程与应用, 2026, 62(13): 312-320.

[3] 何达航, 王昱, 左利云. 边云环境中基于深度强化学习的任务卸载和资源分配方法[J]. 计算机应用研究, 2025, 42(2): 486-493.

[4] Shi Z Y, Xie X Z, Lu H B, et al. DRL-based multidimensional resource management in SWIPT-NOMA-enabled MEC[J]. IEEE Transactions on Wireless Communications, 2024, 23(4): 3252-3266.

[5] Zhang R L, Zhou R T, Wang Y F, et al. Incentive mechanisms for online task offloading with privacy-preserving in UAV-assisted mobile edge computing[J]. IEEE/ACM Transactions on Networking, 2024, 32(3): 2646-2661.

[6] Huang C, Chen G J, Xiao P, et al. Joint offloading and resource allocation for hybrid cloud and edge computing in SAGINs: a decision assisted hybrid action space deep reinforcement learning approach[J]. IEEE Journal on Selected Areas in Communications, 2024, 42(5): 1029-1043.

[7] Gong X W, Chen M Y, Li D S, et al. Delay-optimal distributed computation offloading in wireless edge networks[J]. IEEE/ACM Transactions on Networking, 2024, 32(4): 3376-3391.

作者简介: 赵智宇(1989-), 男, 汉族, 本科, 研究方向为电子信息。