

文章编号: 2095-4980(2024)11-1199-10

基于 5G 的列车云边端协同计算设计与优化

徐建喜¹, 魏思雨², 李宗平³

(1. 国能铁路装备有限责任公司, 北京 100011; 2. 北京交通大学 电子信息工程学院, 北京 100044;
3. 交控科技股份有限公司, 北京 100070)

摘要: 城市轨道交通对于缓解城市交通拥堵具有重要作用, 城轨列车多车协同控制是近年来的研究热点。多车协同计算任务受通信的限制, 存在资源分配平衡差、系统对环境变化响应慢、协同运行能力有限等问题。5G 通信与移动边缘计算(MEC)的结合可有效改进任务处理的实时性与准确性, 提高系统整体性能。本文设计了一种基于 5G 与 MEC 的城轨列车运行控制系统自主协同计算架构, 根据多车协同控制任务的特征, 将多车协同计算卸载中的边缘服务器选择问题建模为多臂匪徒(MAB)学习模型, 并提出一种基于置信区间上限(UCB)算法的求解方案, 使城轨列车多车协同控制系统的整体能耗和时延最小。仿真结果表明, 本文所提出的算法模型在平均奖励、最佳选择概率、平均执行时延、加权总成本等方面具有显著的性能优势。

关键词: 多车协作; 移动边缘(MEC)计算; 5G 网络; 任务卸载; 多臂匪徒(MAB)学习; 置信区间上限(UCB)算法

中图分类号: TN919.3

文献标志码: A

doi: 10.11805/TKYDA2023049

5G-based design and optimization of cloud-edge-train collaborative computing

XU Jianxi¹, WEI Siyu², LI Zongping³

(1.China Energy Railway Equipment Co., Ltd, Beijing 100011, China; 2.School of Electronic Information Engineering, Beijing Jiaotong University, Beijing 100044, China; 3.Traffic Control Technology Co., Ltd., Beijing 100070, China)

Abstract: Urban rail transit plays a significant role in alleviating urban traffic congestion, and the coordinated control of multiple urban rail vehicles has been a research hotspot in recent years. The multi-vehicle coordinated computing task is limited by communication, leading to issues such as poor resource allocation balance, slow system response to environmental changes, and limited cooperative operation capabilities. The integration of 5G communication and Mobile Edge Computing(MEC) can effectively improve the real-time and accuracy of task processing, enhancing the overall system performance. This paper designs an autonomous coordinated computing architecture for urban rail vehicle operation control systems based on 5G and MEC. According to the characteristics of multi-vehicle coordinated control tasks, the problem of edge server selection in multi-vehicle coordinated computing offloading is modeled as a Multi-Armed Bandit(MAB) learning model, and a solution based on the Upper Confidence Bound(UCB) algorithm is proposed to minimize the overall energy consumption and latency of the urban rail vehicle multi-vehicle coordinated control system. Simulation results show that the proposed algorithm model has significant performance advantages in terms of average reward, best selection probability, average execution latency, and weighted total cost.

Keywords: multi-train collaboration; Mobile Edge Computing(MEC); 5G network; task offloading; Multi-Armed Bandits(MAB) learning; Upper Confidence Bound(UCB) algorithm

城市轨道交通具备效率高、运能大等特点, 为居民提供了安全舒适、经济便捷的运输服务^[1-2]。近年来, 随着客流量与线路规模的不断增加, 传统列车运行控制无法满足智能智慧城轨的发展需求, 多车协同控制成为学术界的研究热点之一^[3-4]。多车协同控制系统在动态环境中的自主合作任务受通信的限制, 存在资源分配平衡

收稿日期: 2023-02-28; 修回日期: 2023-05-08

基金项目: 国家自然科学基金资助项目(61973026)

差、系统对环境动态变化反应迟缓、合作运行能力不足等问题^[5]。5G 高带宽、低时延的特性可有效突破多车自主合作的通信瓶颈^[6-7]，提高多车协同运行能力^[8]。但在多车协作中，车-车、车-地的通信需求是多种多样的，如需要传输大量的视频和图像数据包、低延迟和高可靠的列车运行控制指令等。

在任务驱动的多车协作问题中，存在的挑战包括复杂的信息交互逻辑、大量的通信节点与传输数据、节点之间不同的耦合关系、不同的业务信息优先级等^[9]。充分利用 5G 的高带宽和低延迟，最小化响应时间并最大限度地支持协作任务执行是需要解决的重要问题。同时，开放的动态环境导致传输通道复杂多变，存在随机不确定性干扰。因此可靠的通信机制和实时性能也是需要解决的核心问题。

移动边缘计算(MEC)是一种新兴架构，通过在网络边缘部署高性能服务器增强计算能力^[10]。在列车沿线附近部署 MEC 服务器，列车可以通过 5G 网络卸载计算任务到 MEC 服务器。由于边缘服务器位于列车附近，与云服务器相比，数据传输可在更短的时间内完成，传输功率大大降低，从而减少了列车计算任务的卸载延迟和能耗。为充分利用边缘中的计算资源，边缘智能必不可少，列车需要选择合适的边缘服务器，并以自适应和自主的方式完成计算任务。在边缘智能范式中，列车车载设备根据数据计算需求将计算任务转移到边缘服务器。只有从许多可用服务器中选择同时满足延迟和能耗需求的最佳边缘服务器时，边缘智能才能发挥作用。因此，列车需要在不确定的环境中根据延迟和能源消耗的先验信息改进其决策，这是一个非常具有挑战性的问题。同时，在动态环境中，关键变量可能会突然发生变化，使问题更加复杂。卸载中，边缘服务器的选择问题是一种经典的最优决策问题，且列车事先无法获得各个边缘服务器的具体状态参数。针对这种未知感知质量的最优决策问题，多臂匪徒(MAB)学习被证明是一种有效的方法^[11]，MAB 问题是强化学习中一个经典的探索-开发平衡问题^[12]。最初的 MAB 考虑 K 个机器，当赌徒(智能体)拉动杠杆时，每台机器都会提供奖励。奖励遵循每台机器特定的概率分布，但赌徒不知道。通过在当前具有最高预期奖励的机器和其他机器之间多次试验以获得更多信息，赌徒的目标是通过依次选择要玩的机器最大化某些回合内的累积奖励^[13]。本文中，列车被视为赌徒，边缘服务器被视为多臂机器。列车最初在没有任何先验信息的情况下选择边缘服务器，随后在卸载过程中收集时间和能源成本的信息，并在以后的迭代中相应地改进对服务器的选择，即实现一个预算受限的 MAB 问题。解决 MAB 问题的策略主要有 Epsilon-Greedy 算法、Boltzmann 策略算法与置信区间上限(UCB)策略算法^[14-16]，通过对比 3 种策略算法的收益变化与摇臂决策概率变化情况^[17]，UCB 策略算法更适用于多臂匪徒问题，能够快速选取最优决策并获得更高的决策收益。

针对基于 5G 的城轨列车云边端协同计算问题，本文设计了一种用于多车协作的集中-分布式架构。该架构充分利用云、边、端协同，为多车协同提供计算资源，同时全面优化计算延迟和能耗。随后，将多车协同控制任务卸载与资源分配的问题建模为一个预算受限的 MAB 问题，并采用 UCB 方法求解。

1 系统模型

1.1 系统架构

在复杂多变的动态环境中，列车需要完成不同的计算密集型任务。单个计算任务可以在本地执行，也可以卸载到附近的边缘服务器。但在一堆计算密集型任务的高数据传输密度场景下，边缘服务器的计算资源可能会耗尽，无法满足所有卸载需求。因此，对于列车车载设备，选择合适的服务器降低卸载过程中的延迟和能耗尤为重要。本文提出了一种基于 5G 与边缘计算的集中-分布式云边端架构，如图 1 所示。系统从上到下由云服务层、边缘层和车载层组成。云服务层负责管理所有列车，通过无线网络与边缘计算通信，并将任务和指令发送给边缘节点；边缘层由多个边缘服务器组成，负责响应车载设备的计算请求；车载层主要是指列车及其携带的处理器和传感器。

如图 2 所示，列车可以通过云边端协作架构协同工作。车载设备通过简短的信息共享获取周围列车与边缘服务器的信息，并使用这些信息决定自己的决策行为。对于计算密集型任务，列车可以通过 5G 通信基站将任务卸载到边缘服务器，同时还可以选择将棘手的任务上传到更强大的云计算中心。MEC 通过为每个列车的任务合理分配计算资源，提高了整个系统的协作效率。

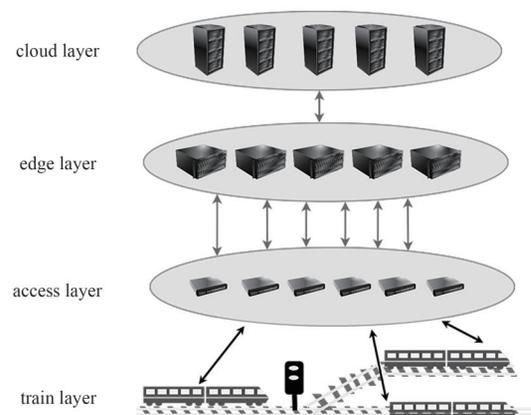


Fig.1 Centralized-distributed system architecture based on 5G and edge computing

图 1 基于 5G 与边缘计算的集中-分布式系统架构

在这种架构下，当列车面临大数据包的视频图像分析和高时延敏感性计算等问题时，可通过卸载任务和资源分配进行高效处理。本文将列车任务卸载和资源问题划分为 3 个部分：网络模型、任务模型和计算模型。

1.2 网络模型

在多车协同系统中，列车车载计算机的计算能力是有限的。假设列车集合 $N=\{1,2,\dots,N\}$ 处于边缘服务器集合 $D=\{1,2,\dots,D\}$ 的服务范围内，任意有卸载需求的列车可以选择任意边缘服务器运行计算密集型的任务。边缘服务器部署在 5G 基站附近，并与 5G 基站连接^[18]。

将 W 定义为 5G 基站的网络带宽。由于通道是正交的，因此不考虑间隙干扰。当列车将任务发送到边缘服务器节点时，可实现的数据上传速率为：

$$R_d = W \log_2 \left(1 + \frac{p_d g_d}{W N_d} \right) \quad (1)$$

式中： p_d 为车载设备的发射功率； g_d 为车载设备与边缘服务器之间的信道增益； N_d 为信道中高斯白噪声的功率谱密度。

1.3 任务模型

列车车载设备是一种可与边缘节点通信的信息传感设备。MEC 节点是具有通信和计算能力的服务器，位于列车附近和核心网络的边缘。每个车载设备可从 D 个服务器中选择一个，通过带宽 W 建立通信并卸载计算密集型任务。为表示列车的卸载决策，本文定义了二元指标 $\alpha_n = \{0, 1\}$ ，在本地处理任务时 $\alpha_n = 0$ ，在边缘服务器中处理时 $\alpha_n = 1$ 。假设列车的计算密集型任务为 $R_n = (D_n, B_n)$ ，其中， B_n 表示计算任务的数据大小， D_n 表示处理任务所需的 CPU 周期，列车的每个任务不能分成多个子任务，这意味着任务要么全部在本地处理，要么全部卸载到 MEC 服务器处理。

1.4 计算模型

1.4.1 本地计算模型

当列车 n 决定在本地执行任务 R_n 时，定义 T_n^1 为本地执行延迟时间， f_n^1 为车载设备的计算资源， D_n 为任务大小。不同的列车可能有不同的计算能力，因此任务 R_n 的本地处理时间 T_n^1 表示为：

$$T_n^1 = \frac{D_n}{f_n^1} \quad (2)$$

定义 E_n^1 为处理任务 R_n 所消耗的能源：

$$E_n^1 = Z_n D_n \quad (3)$$

式中 Z_n 为车载设备上计算单元执行任务 R_n 的每个计算周期的能耗，根据文献[19]，设 $Z_n = 10^{-27} (f_n^1)^2$ 。

定义 C_n^1 为在本地车载设备上处理任务 R_n 的总成本：

$$C_n^1 = I^t T_n^1 + I^e E_n^1 \quad (4)$$

式中 I^t 和 I^e 表示权重，满足 $0 \leq I^t \leq 1$ ， $0 \leq I^e \leq 1$ ， $I^t + I^e = 1$ 。在仿真过程中，保持权重不变。

1.4.2 MEC 计算模型

当列车 n 选择在 MEC 服务器上执行任务 R_n 时，可以将处理过程分为 3 个步骤：首先，列车将数据上传至 5G 基站，并经 5G 基站传递至 MEC 服务器；然后，MEC 为任务 R_n 分配计算资源，并处理数据以获得最终结果；最后，MEC 通过 5G 基站将结果发回列车 n 。

定义 $T_{n,t}^o$ 为卸载时间：

$$T_{n,t}^o = \frac{B_n}{R_d} \quad (5)$$

这一步消耗的能量为：

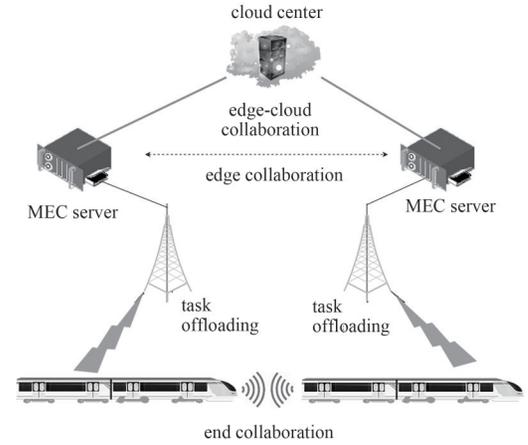


Fig.2 Structure of cloud-edge-end collaboration centralized-distributed system

图2 云-边-端协作集中-分布式系统结构

$$E_{n,t}^{\circ} = P_n T_{n,t}^{\circ} = \frac{P_n B_n}{R_d} \quad (6)$$

式中 P_n 为列车 n 的功率消耗。

定义 $T_{n,p}^{\circ}$ 为 MEC 服务器处理任务 R_n 所需的时间:

$$T_{n,p}^{\circ} = \frac{D_n}{f_n} \quad (7)$$

则列车 n 消耗的能源可表示为:

$$E_{n,p}^{\circ} = P_n T_{n,p}^{\circ} = \frac{P_n D_n}{f_n} \quad (8)$$

第 3 个过程中返回结果的数据通常非常小, 而 5G 网络具有非常快的网络下行速率, 因此本文中忽略返回数据的时延与能耗值。

将式(5)和式(7)相加, 可得运行任务 R_n 的整体时延为:

$$T_n^{\circ} = \frac{B_n}{R_d} + \frac{D_n}{f_n} \quad (9)$$

将式(6)和式(8)相加, 可得运行任务 R_n 的整体能耗为:

$$E_n^{\circ} = \frac{P_n B_n}{R_d} + \frac{P_n D_n}{f_n} \quad (10)$$

将时间成本(式(9))和能源成本(式(10))相加, 得到 MEC 的总成本为:

$$C_n^{\circ} = I^t T_n^{\circ} + I^e E_n^{\circ} \quad (11)$$

则任务 R_n 的总成本可表示为:

$$C_{\text{all}} = \sum_{n=1}^N (1 - \alpha_n) C_n^l + \alpha_n C_n^{\circ} \quad (12)$$

式中 α_n 为折扣因子。

1.5 问题建模

本文制定了边缘服务器选择问题, 将卸载与运算受限的 MAB 问题结合。模型的目标是: 在预算限制下, 赌徒(车载设备)应该选择哪台机器(边缘服务器)最大化每个赌注的利润? 如何“组合手臂(选择边缘服务器)”以在有限的回合中实现利润最大化? 在 MAB 框架下, 本文制定一个策略为任务选择最合适的边缘服务器。

1.5.1 奖励和成本模型

每个列车在总时延或能耗方面都有一定的服务限制, 因此, 根据上述考虑因素选择最佳边缘服务器至关重要, 但在不确定和动态的网络环境中, 选择最佳的边缘服务器并非易事。任务卸载旨在最大化奖励并最小化卸载期间的成本。当列车决定卸载时, 根据式(9), 在时延 T_n° 之后, 任务才会完成并返回。卸载的奖励和成本定义如下:

1) 奖励: 首先, 根据时间延迟分析执行多个任务所需的总时间。任务的处理方法根据不同的计算任务进行变化, 有些计算任务很小, 可由列车本身在本地处理, 因此可以很容易获得所需时间 T_n^l ; 如果任务被卸载到边缘服务器, 则所需时间为 T_n° 。假设列车车载设备到边缘服务器资源的网络带宽非常宽, 卸载可以提高应用程序的相关性能, 减少执行任务的时间, 并降低总计算和通信成本。卸载节省的时间为 $T_n^l - T_n^{\circ}$, 如果执行任务所需的时间减少, 则节省的时间一定大于 0。为使列车获得更高的服务质量, 本文定义选择边缘服务器卸载的奖励为:

$$r_c = \begin{cases} 0, & T_n^{\circ} > \theta \\ 1, & T_n^{\circ} \leq \theta \end{cases} \quad (13)$$

式中 θ 为门限值。延迟时间 T_n° 必须保持在预先指定的 θ 值以下 ($T_n^{\circ} < \theta$), 用户服务质量的要求可以满足, 否则满足不了。

2) 成本: 能源消耗是边缘计算的重要限制因素, 因此, 消耗的能量是任务成本的重要因素。由上文可得,

选择将任务卸载至边缘服务器时，加权了时延和能耗的总成本函数可表示为式(11)。

1.5.2 预算受限的 MAB 任务卸载问题

考虑集合为 $D = \{1, 2, \dots, K\}$ 的边缘服务器作为 MAB 的一套多臂机器。在一个回合中，智能体拉动手臂 $k \in D$ ，即选择边缘服务器 k ，奖励为 r_{is}^k ，成本为 c_{is}^k (下标 is 表示理想情况)， $M_k(n)$ 为边缘服务器 k 在总选择为 n 次时被选择的次数。根据当前选择策略，在进行 n 次选择之后可得成本为：

$$\lambda = nr^* - \sum_{k=1}^K E[M_k(n)]r_k \quad (14)$$

式中： r^* 为最优边缘服务器的平均值； r_k 为选择边缘服务器之后产生的平均回报。产生成本的原因是使用不同的选择策略，但并不总是能够每次都选择到最优的边缘服务器。奖励变量是有上限的，因此存在一个常数 r_{is}^{\max} 且 $r_{is}^k \leq r_{is}^{\max}$ 。同样，车载设备和边缘服务器之间的卸载成本也是一个随机变量，成本变量具有下限，即 $c_{is}^k \geq c_{is}^{\min}$ 。智能体必须确保成本 c_{is}^k 在下一轮迭代后满足给定的预算限制 Y 。假设最理想状态持续到最后一次迭代，智能体会得到累计最大的回报 r_{is} ，记为 Q ：

$$\max \sum_{is}^Q r_{is}^k, \text{ s.t. } \sum_{is}^Q c_{is}^k \leq Y \quad (15)$$

这是在每一轮中手臂 k 被选择后的成本和回报。此时选择最优边缘服务器的平均回报为：

$$r^* = \max(r_{is}^k) \quad (16)$$

车载设备的目标是找到最优的手臂 k ，即最合适的边缘服务器，并将遗憾值降至最低。使用某种算法策略记录 T 轮后的遗憾值，使其最小：

$$R_c = Tr^* - \sum_{is}^T r_{is}^k \quad (17)$$

事实上，并不知道奖励和成本相关的随机变量产生的结果， Q 也是一个随机变量，与奖励或成本之和的随机变量有关。本文还考虑了拉动一次手臂 k 时的奖励和成本的加权总和，以评估选择手臂 k 的好处，并将其定义为效用函数。当列车选择边缘服务器 N 卸载计算任务时，将其看作智能体拉动了一次手臂 $k (k \in N)$ ，此时考虑实际效用函数：

$$u_{rs} = \frac{r_{rs}}{c_{rs}} \quad (18)$$

式中下标 rs 表示真实情况。

在不考虑实际复杂环境的理想条件下，定义最优效用函数为：

$$u_{is} = \frac{r_{is}}{c_{is}} \quad (19)$$

将列车选择将计算任务卸载到边缘服务器 N 时的遗憾函数定义为：

$$R_c(R) = \frac{r_{is}}{c_{is}} - \frac{r_{rs}}{c_{rs}} \quad (20)$$

因此，累计的遗憾值为：

$$R_Q = \sum_{is}^Q \left(\frac{r_{is}}{c_{is}} - \frac{r_{rs}}{c_{rs}} \right) \quad (21)$$

在实际的多车协作环境中，列车的最终目标是使预期遗憾函数最小： $\min E[R_Q]$ 。

基于上述工作，提出一种新的算法。首先，将平均奖励和平均成本定义为：

$$\begin{cases} \bar{r}_T(\tau, k) = \frac{\sum_{i=\tau-1}^T r_{k,i} \Pi(I_i=k)}{L_T(\tau, k)} \\ \bar{c}_T(\tau, k) = \frac{\sum_{i=\tau-1}^T c_{k,i} \Pi(I_i=k)}{L_T(\tau, k)} \end{cases} \quad (22)$$

式中:

$$L_T(\tau, k) = \sum_{i=T-\tau+1}^T \Pi(I_i=k) \quad (23)$$

式中: T 表示总时间步数; τ 表示时间窗口的长度; i 表示时刻; I_i 表示 i 时发生的事件。

此外, 车载信息卸载到边缘服务器进行计算累计消耗的计算能耗为:

$$E_T(\tau, k) = \frac{\left(\frac{r_{\max}^k}{c_{\min}} + 1\right) r_{\max}^k \sqrt{\frac{\gamma \log(\min(T, \tau))}{L_T(\tau, k)}}}{c_{\min}^k - r_{\max}^k \sqrt{\frac{\gamma \log(\min(T, \tau))}{L_T(\tau, k)}}} \quad (24)$$

式中 τ 和 γ 为可调节的参数。

车载设备的成本不仅是结合延迟和能源消耗的总加权成本, 还需考虑边缘服务器和车载设备之间通信的路径损失^[20]。因此, 边缘服务器完成整个任务的总成本为加权和(C_n^o)与路径损耗($P(d)$)的总和, 即

$$c_{rs} = C_n^o + P(d) = I^T T_n^o + I^c E_n^o + 20 \log\left(\frac{4\pi f_c d}{c}\right) \quad (25)$$

式中: d 为列车与边缘服务器之间的直线距离; c 为光速; f_c 为载体频率。

列车可以选择本地计算或将任务卸载至边缘服务器, 并考虑多车协作环境中分配的网络资源和计算资源。本文的优化目标是 minimized 遗憾函数。因此, 本文制定的 MAB 问题如下:

$$\begin{cases} \min_{E[R_O]} \sum_{n=1}^N \frac{r_{is} - r_{rs}}{c_{is}} \\ \text{s.t. } C_1: r_{is}^k \leq r_{is}^{\max}, \forall k \in N, n \\ C_2: c_{is}^{\min} \leq c_{is}^k \leq Y, \forall k \in N, n \\ C_3: I^t \geq 0, I^c \leq 1 \\ C_4: \min[R_O] \end{cases} \quad (26)$$

约束条件必须确保满足最大延迟约束和服务器资源预算, 即必须确保满足车载设备的最大奖励和最小成本的约束(C_1 和 C_2)、最终总加权和最小遗憾函数(C_4)。这个优化问题的一个挑战是, 随着列车数量的增加, 问题的规模和复杂性也将迅速增加。为此, 本文提出一种新的基于强化学习的算法。

2 基于强化学习的求解方案

当一个机器设备选择最优的边缘服务器进行任务卸载时, 是应该探索所有的边缘服务器, 还是直接选择目前提供最大累积奖励的服务器, 是 MAB 问题的核心。这是“探索”(估计每支手臂的奖励价值)和“开发”(选择当前性能的最佳手臂)之间的矛盾。因为游戏数量有限, 因此尝试总数是固定的且有限的, 更多“探索”将不可避免地减少“开发”的数量, 反之亦然。因此, 有必要在两者之间进行有效的平衡并做出决定, 以获得最大的累积回报。考虑到机器人协作环境充满了随机条件, 且缺乏对这些条件的先验知识, 因此本文采用基于非确定条件下的乐观原则的 UCB 算法探索最优边缘服务器。

2.1 传统 UCB 算法

UCB 算法有助于根据置信区间选择机器手臂^[21]。置信区间可简单地理解为不确定性的程度。区间越宽, 不确定性越大。随着实验的进行, 计算并更新了多臂匪徒机器的平均手臂回报。执行实验的次数越多, 根据中心极限定理得出的统计概率就越接近真实概率。即随着试验次数的增加, 置信区间会变窄, 这意味着置信区间会越来越准确, 上限将逐渐接近真实值。最后, 选择带有最高置信区间的手臂, 即边缘服务器。算法流程如图 3 所示。

UCB 算法的基本流程为: a) 选择平均奖励较大的手臂及其 UCB; b) 拉动手臂以获得奖励; c) 更新手臂的奖励和置信区间。该算法的优点是在算法决策中考虑了统计值的不确定性, 无需预设参数。同时, UCB 算法不受手臂返回值随机分布的影响。

2.2 用于选择边缘服务器的 UCB1-ESSS 算法

UCB1 算法是一个简单且扩展的 UCB 公式^[22]。在 UCB 算法的基础上，UCB1 算法不仅注重回报和成本，还注重探索次数。基于 UCB1 算法和列车车载设备的卸载需求，本文提出一种新的基于 UCB1 的边缘服务器选择策略算法，即 UCB1-ESSS(UCB1 with Exploration via Sampled Successive Sharpening) 算法，以解决相关的 MAB 优化问题。列车车载设备分析历史选择的记录和相关奖励、成本值，以确定在下一轮中选择哪个边缘服务器；然后更新每个边缘服务器的置信区间、选择的次数；最后选择遗憾值最低、满足卸载要求的边缘服务器作为任务卸载的最优边缘服务器。

根据式(27)~(28)，列车车载设备选择对应的边缘服务器：

$$\text{arm}(T) = \arg \max_T \left[\bar{r}_k + \sqrt{\frac{2\log(T)}{M(k)}} \right] \quad (27)$$

式中： T 为总回合数； $\text{arm}(T)$ 为第 T 轮中选择的边缘服务器； \bar{r}_k 为前所选的卸载服务器在前面实验中的平均奖励值； $M(k)$ 为边缘服务器 k 被选择的次数， $\sqrt{\frac{2\log(T)}{M(k)}}$ 称为奖金。 $\sqrt{\frac{2\log(T)}{M(k)}}$ 本质是均值的标准差，反映了服务器选择效果的不确定性，需继续探索未探索的边缘服务器，直至达到所需的置信区间上限。

$$\text{arm}(T) = \arg \max_T \left[\bar{r}_k + \beta \sqrt{\frac{(1+\alpha)\log[T/\omega(M(k))]}{2\omega(M(k))}} \right] \quad (28)$$

$$\omega(n) = (1+\alpha)^n \quad (29)$$

式中： α 用于控制探索部分的影响程度，是一个处于(0,1)范围的参数； β 为实际奖励的标准差，表示给予探索的权重，由于奖励是一个独立分布的随机变量，因此 β 的值只能通过实际测试得出。具体过程如算法 1 所示。

算法 1：基于 UCB1-ESSS 的边缘计算服务器选择策略

输入： K 个手臂，迭代次数 $T \geq K$

输入：参数 τ 、 γ 、 r_{\max}^k 、 c_{\min}^k

评估参数

for $k = 1, 2, \dots, N$ do

 拉下手臂 k

 记录回报 r_{is}^k 和成本 c_{is}^k

end for

while $\sum_{\text{is}}^Q c_{\text{is}}^k \leq Y$ do

 拉动手臂 k ，当满足

$$k = \arg \max \frac{\bar{r}_T(\tau, k)}{\bar{c}_T(\tau, k)} + E_T(\tau, k)$$

 记录回报 r_{is}^k 和成本 c_{is}^k

 拉动手臂 k ，当满足

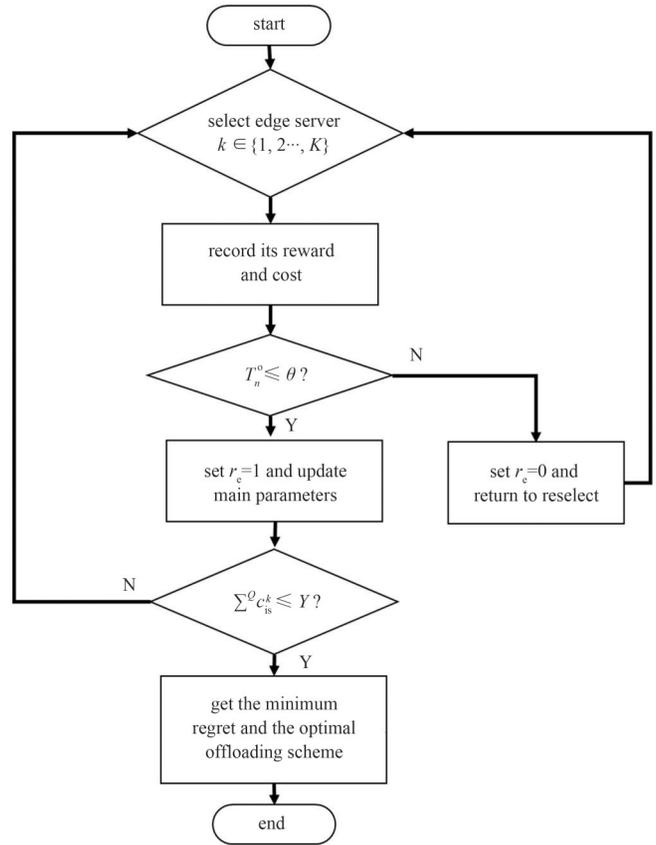


Fig.3 Flow chart of traditional UCB algorithm
图3 传统 UCB 算法流程图

$$\text{arm}(T) = \arg \max_T \left[\bar{r}_k + \beta \sqrt{\frac{(1 + \alpha) \log [T/\omega(M(k))]}{2\omega(M(k))}} \right]$$

记录遗憾值

end while

3 仿真结果

假设模拟环境是一个由多辆列车和 3 个边缘服务器组成的小型集成网络。首先将所提出的 UCB1-ESSS 算法与传统的 UCB 算法和 Epsilon-Greedy 算法进行比较。Epsilon-Greedy 算法也是一种基于 MAB 模型的策略，其基本思想描述如下：首先选择一个非常小的数字 $\varepsilon \in (0, 1)$ 。每一轮，智能体根据现有的历史经验，以概率 $1 - \varepsilon$ 选择平均回报最大的手臂，或以概率 ε 随机选择任意一条手臂。在每一轮中，默认情况是选择迄今为止平均奖励最好的手臂(开发)，偶尔选择其他手臂之一(探索)。探索是完全随机的。如果 ε 随着时间的推移动态修改，则可以获得更好的性能^[23]。

根据 T 轮的平均奖励和选择最佳手臂的概率比较 3 种算法的性能。MAB 问题的数学描述是通过选择最佳手臂获得最高的平均回报，并采用算法在 T 轮的最小化遗憾值。本文中最好的手臂是卸载任务时延迟和能耗最小的边缘服务器。

假设 5G 基站带宽为 $W = 20$ MHz，5G 基站连接到部署在其旁边的 MEC，5G 基站与 MEC 的传输延迟可忽略。MEC 服务器的计算资源为 $F(f_n) = 25$ GHz/s，车载设备的传输功率为 $p_d = 100$ mW。卸载数据大小 B_n 设置为 (300, 500) kbits 之间的随机值，CPU 周期数 D_n 为介于 (900, 1 100) 之间的随机值，同时设置 $\beta = 1$ ， $\varepsilon = 0.1$ 。

图 4 为 3 种算法实现的平均奖励的对比。可以看出，UCB1-ESSS 算法在学习的前 1 000 次迭代中，比其他两种算法的平均奖励评价更高，可以更快地挑选最优边缘服务器。需要注意的是，在所提出的 UCB1-ESSS 算法中，为提高每台边缘服务器的置信度，也会选择平均奖励较低的服务器，因此收敛速度较慢，但平均奖励高于其他两种算法。

图 5 为 3 种算法选择最佳手臂的概率对比。可以看出，一开始，所有边缘服务器的置信度都很低，每个服务器都会被选中几次，此时，选择最优边缘服务器的概率并不高。列车多次选择后，每个服务器的置信度都相当高，此时，列车总会选择最优的边缘服务器。可以清楚地看到，UCB1-ESSS 算法的波动性高于其他两种算法，在仿真结束时，UCB1-ESSS 算法选择最优边缘服务器的概率已经超过了 UCB 算法和 Epsilon-Greedy 算法。

图 6 为 UCB1-ESSS、UCB 和 Epsilon-Greedy 算法在平均延迟方面的性能比较。同时比较了任务完全在本地计算的平均时延。可以看出，与本地处理相比，计算任务数量越大，平均延迟越大。这是因为更多的任务会导致对有限计算资源的更激烈的竞争，任务无法及时完成，导致延迟增加。所提算法的性能与传统 UCB 算法的性能非常接近，证明了所提出的算法表现更好，因为它在相同甚至略低的延迟下获得了更高的奖励。

图 7 为系统任务处理的加权总成本随着计算任务数量增加的变化情况。所提算法的加权总成本最低，UCB 算法的性能稍差；同时，两种方法的性能相对稳定。当计算任务数量较少时，3 种算法选择的服务器差别不大，当任务数量增加时，总开销增加，因为计算资源有限，这反映了所选边缘服务器能否在保证性能的同时提供充足的计算资源。所提算法能够满足需求。

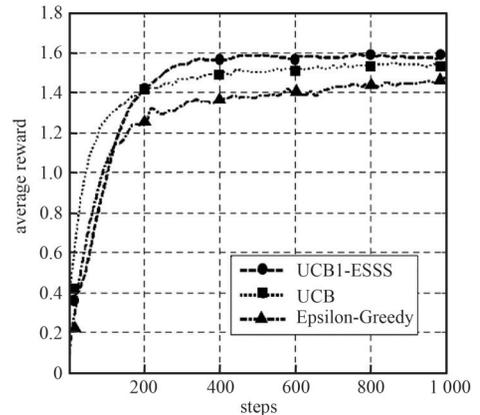


Fig.4 Comparison of the average rewards of the three algorithms

图 4 3 种算法的平均奖励对比

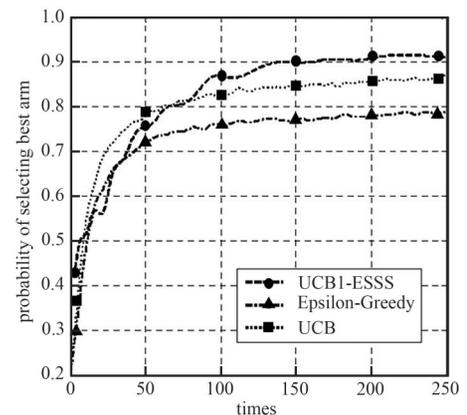


Fig.5 Comparison of the probability of choosing the best arm among the three algorithms

图 5 3 种算法中选择最佳手臂概率的对比

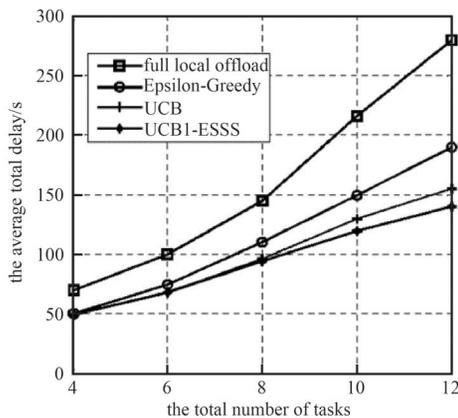


Fig.6 Comparison of average total delay
图6 平均总时延对比

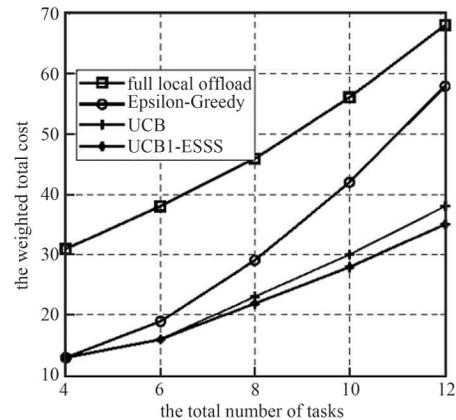


Fig.7 Comparison of weighted total cost
图7 加权总成本对比

4 结论

本文提出了一种基于 5G 和边缘计算的多列车自主协作方法框架。在此框架下，提出了一种基于传统 UCB 算法的新型 UCB1-ESSS 算法，以确定最适合卸载任务的边缘服务器。仿真结果表明，UCB1-ESSS 算法在各种条件下均优于传统 UCB 算法和 Epsilon-Greedy 算法，实现了更低的延迟和加权总成本。为更好地理解列车协作通信网络资源的优化利用，未来将继续致力于边缘智能的研究，优化集成网络资源，并解决由此产生的大规模调度问题。

参考文献：

- [1] 陈卓. 城市轨道交通通信系统运行现状与发展趋势探究[J]. 现代工业经济和信息化, 2022,12(9):254-256. (CHEN Zhuo. Exploring the current situation and development trend of urban rail transit communication system operation[J]. Modern Industrial Economy and Informationization, 2022,12(9):254-256.) doi:10.16525/j.cnki.14-1362/n.2022.09.105.
- [2] 龙俊仁,邵源,牛雄. 城市轨道交通绿色发展问题思考[J]. 城市交通, 2022,20(5):60-66. (LONG Junren,SHAO Yuan,NIU Xiong. Thoughts on green development of urban rail transit[J]. Urban Transport of China, 2022,20(5):60-66.) doi:10.13813/j.cnki.11-5141/u.2022.0507.
- [3] 杨舜尧. 基于多车协同的城市轨道交通节能优化方法[D]. 北京:北京交通大学, 2022. (YANG Shun Yao. Energy-saving optimization method for urban rail transit based on multi-vehicle coordination[D]. Beijing:Beijing Jiaotong University, 2022.)
- [4] 李波. 基于数据驱动的自适应控制器设计及其在多车协同控制中的应用[D]. 上海:上海交通大学, 2020. (LI Bo. Data-driven adaptive controller design and its application in multi-vehicle cooperative control[D]. Shanghai,China:Shanghai Jiaotong University, 2020.) doi:10.27307/d.cnki.gsytu.2020.001921.
- [5] YANG Songpo,LIAO Feixiong,WU Jianjun,et al. A bi-objective timetable optimization model incorporating energy allocation and passenger assignment in an energy-regenerative metro system[J]. Transportation Research Part B:Methodological, 2020 (133):85-113. doi:10.1016/j.trb.2020.01.001.
- [6] 毕健有,王文成,王家旭. 5G 技术在工业互联网中的应用研究[J]. 数字通信世界, 2022(6):150-152. (BI Jianyou,WANG Wencheng,WANG Jiayu. Research on the application of 5G technology in industrial Internet[J]. Digital Communication World, 2022(6):150-152.) doi:10.3969/J.ISSN.1672-7274.2022.06.049.
- [7] 马占军. 基于 5G 网络的工业物联网架构[J]. 中国新通信, 2020,22(18):18-19. (MA Zhanjun. Architecture of industrial Internet of things based on 5G network[J]. China New Telecommunications, 2020,22(18):18-19.) doi:10.3969/j.issn.1673-4866.2020.18.010.
- [8] 李沁颖,易豪,孙翊馨. 面向协同驾驶的基于移动边缘计算的 5G 智能网联车辆服务平台[J]. 南方农机, 2022,53(11):24-26, 32. (LI Qinying,YI Hao,SUN Yixin. A 5G intelligent connected vehicle service platform based on mobile edge computing for collaborative driving[J]. South Agricultural Machinery, 2022,53(11):24-26,32.) doi:10.3969/j.issn.1672-3872.2022.11.007.
- [9] ABBAS N,ZHANG Y,TAHERKORDI A,et al. Mobile edge computing:a survey[J]. IEEE Internet of Things Journal, 2018,5(1): 450-465. doi:10.1109/JIOT.2017.2750180.
- [10] 陆鹏威,颜子彦,张伟,等. 一种面向电磁识别模型的分散计算方法[J]. 太赫兹科学与电子信息学报, 2022,20(1):22-28. (LU Pengwei,YAN Ziyuan,ZHANG Wei,et al. Decentralized calculation of neural network model for electromagnetic object detection[J].

- Journal of Terahertz Science and Electronic Information Technology, 2022,20(1):22–28.) doi:10.11805/TKYDA2021153.
- [11] ALMEIDA C, GONCALVES R, VENSKE S, et al. Multi-armed bandit based hyper-heuristics for the permutation flow shop problem[C]// The 7th Brazilian Conference on Intelligent Systems. Sao Paulo, Brazil: IEEE, 2018:139–144.
- [12] KAWATA R, FUJITA K. Meta-strategy based on multi-armed bandit approach for multi-time negotiation[J]. IEICE Transactions on Information and Systems, 2020,103(12):2540–2548. doi:10.1587/transinf.2020SAP0003.
- [13] ALMEIDA C P, GONÇALVES R A, VENSKE S, et al. Hyper-heuristics using multi-armed bandit models for multi-objective optimization[J]. Applied Soft Computing, 2020(95):106520. doi:10.1016/j.asoc.2020.106520.
- [14] SINGH S, TRIVEDI A. Anti-jamming in cognitive radio networks using reinforcement learning algorithms[C]// 2012 the Ninth International Conference on Wireless and Optical Communications Networks(WOCN). Indore, India: IEEE, 2012: 1–5. doi:10.1109/WOCN.2012.6331885.
- [15] AUER P. Using upper confidence bounds for online learning[C]// The 41st Annual Symposium on Foundations of Computer Science. Redondo Beach, CA, USA: IEEE, 2000:270–279. doi:10.1109/SFCS.2000.892116.
- [16] ASADI K, LITTMAN M. A new Softmax operator for reinforcement learning[DB/OL]. (2016–12–16)[2023–02–28]. <https://arxiv.org/abs/1612.05628v3>.
- [17] 毛少卿. 基于强化学习的智能干扰决策方法研究[D]. 哈尔滨: 哈尔滨工业大学, 2022. (MAO Shaoqing. Research on intelligent interference decision-making method based on reinforcement learning[D]. Harbin, China: Harbin Institute of Technology, 2022.)
- [18] 黄冬艳, 付中卫, 王波. 计算资源受限的移动边缘计算服务器收益优化策略[J]. 计算机应用, 2020,40(3):765–769. (HUANG Dongyan, FU Zhongwei, WANG Bo. Revenue maximization strategy for mobile-edge computing server with limited computing resources[J]. Journal of Computer Applications, 2020,40(3):765–769.) doi:10.11772/j.issn.1001–9081.2019081351.
- [19] DENG Shuiguang, ZHAO Hailiang, FANG Weijia, et al. Edge intelligence: the confluence of edge computing and artificial intelligence[J]. IEEE Internet of Things Journal, 2020,7(8):7457–7469. doi:10.1109/JIOT.2020.2984887.
- [20] 王鼎, 门昌骞, 王文剑. 一种核的上下文多臂赌博机推荐算法[J]. 智能系统学报, 2022,17(3):625–633. (WANG Ding, MEN Changqian, WANG Wenjian. A kernel contextual bandit recommendation algorithm[J]. CAAI Transactions on Intelligent Systems, 2022,17(3):625–633.) doi:10.11992/tis.202105039.
- [21] 吴卿源, 谭晓阳. 基于UCB算法的交替深度Q网络[J]. 南京师范大学学报(工程技术版), 2022,22(1):24–29. (WU Qingyuan, TAN Xiaoyang. Alternated deep Q network based on upper confidence bound[J]. Journal of Nanjing Normal University (Engineering and Technology Edition), 2022,22(1):24–29.) doi:10.3969/j.issn.1672–1292.2022.01.004.
- [22] PATIL V, GHALME G, NAIR V, et al. Achieving fairness in the stochastic multi-armed bandit problem[J]. The Journal of Machine Learning Research, 2021,22(1):174. doi:10.5555/3546258.3546432.
- [23] COMPARE M, BARALDI P, MARELLI P, et al. Partially observable Markov decision processes for optimal operations of gas transmission networks[J]. Reliability Engineering & System Safety, 2020(199):106893. doi:10.1016/j.ress.2020.106893.

作者简介:

徐建喜(1989–), 男, 本科, 工程师, 主要研究方向为重载铁路智能运维 .email:11658299@ceic.com.

魏思雨(1999–), 女, 在读硕士研究生, 主要研究方向为列车运行控制系统.

李宗平(1984–), 男, 本科, 工程师, 主要研究方向为列车数据通信系统.