计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2024,33(7):26−38 [doi: 10.15888/j.cnki.csa.009547]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于学习博弈和契约论的分层联邦学习隐私保护激励机制①

宋　彪, 薛　涛, 刘俊华

(西安工程大学 计算机科学学院, 西安 710600)
通信作者: 薛　涛, E-mail: xuetao@xpu.edu.cn

摘　要: 分层联邦学习 (hierarchical federated learning, HFL) 旨在通过多层架构的协作学习, 同时保护隐私和优化模型性能. 但其效果需依赖于针对参与各方的有效激励机制及应对信息不对称的策略. 为了解决上述问题, 本文提出一种保护终端设备、边缘服务器及云服务器隐私的分层激励机制. 在边端层, 边缘服务器作为中介应用多维合约理论设计不同类型的契约项, 促使终端设备在不泄露数据采集、模型训练以及模型传输成本的情况下, 使用本地数据参与 HFL. 在云边层, 云服务器与边缘服务器间关于单位数据奖励和数据量的关系通过 Stackelberg 博弈进行建模, 在不泄露边缘服务器单位利润的情况下, 进一步将其转化为马尔可夫过程, 并采用保护隐私的多智能体深度强化学习 (multi-agent deep reinforcement learning, MADRL) 方法逐渐接近斯塔克伯格均衡 (Stackelberg equilibrium, SE). 实验结果表明, 本文提出的分层激励机制在性能上优于基线方法, 云服务器的收益提升了接近 11%, 单位成本获取增益提升接近 18 倍.

关键词: 分层联邦学习; 博弈论; 多维契约理论; 多智能体深度强化学习; 激励机制

引用格式: 宋彪,薛涛,刘俊华.基于学习博弈和契约论的分层联邦学习隐私保护激励机制.计算机系统应用,2024,33(7):26–38. http://www.c-s-a.org.cn/1003-3254/9547.html

Privacy-preserving Incentive Mechanism for Hierarchical Federated Learning Combining Learning Game and Contract Theory

SONG Biao, XUE Tao, LIU Jun-Hua

(School of Computer Science, Xi'an Polytechnic University, Xi'an 710600, China)

Abstract: Hierarchical federated learning (HFL) aims to optimize model performance and maintain data privacy through multi-layered collaborative learning. However, its effectiveness relies on effective incentive mechanisms for participating parties and strategies to address information asymmetry. To address these issues, this study proposes a layered incentive mechanism for protecting the privacy of end devices, edge servers, and cloud servers. At the edge-device layer, edge servers act as intermediaries, using the multi-dimensional contract theory to design a variety of contract items. This encourages end devices to participate in HFL using local data without disclosing the costs of data collection, model training, and model transmission. At the cloud-edge layer, the Stackelberg game models the relationship between unit data reward and data size between a cloud server and edge servers and subsequently transforms it into a Markov process, all while maintaining the confidentiality of the edge servers' unit profit. Then, multi-agent deep reinforcement learning (MADRL) is used to incrementally approach the Stackelberg equilibrium (SE) while ensuring privacy. Experimental results indicate that the proposed incentive mechanism outperforms traditional approaches, yielding an almost 11% increase in cloud server revenue and an approximately 18 times improvement in the cost-effectiveness gained.

① 基金项目: 国家自然科学基金青年科学基金 (62202366)
收稿时间: 2023-12-26; 修改时间: 2024-01-23; 采用时间: 2024-02-26; csa 在线出版时间: 2024-05-31
CNKI 网络首发时间: 2024-06-04

Key words: hierarchical federated learning (HFL); game theory; multi-dimensional contract theory; multi-agent deep reinforcement learning; incentive mechanism

1 引言

随着移动设备广泛使用和私有数据急剧增长, 依托大数据的机器学习技术实现了在云端的集中式训练. 但由于本地终端设备数据涉及敏感的个人信息, 集中式训练易引发隐私泄露风险, 使用户不愿上传数据至云端[1,2]. 为解决集中式训练存在的数据安全问题, 谷歌提出了一种新型的分布式机器学习-联邦学习 (federated learning, FL)[3]. FL 通过去中心化方式, 实现了数据在各终端设备上的本地存储和模型参数更新, 从而实现模型的协同训练. 然而, 在 FL 中, 为确保模型训练达到既定精度, 大量模型参数需跨复杂网络多轮传输, 导致网络拥塞和通信故障等问题[4]. 针对上述题, 分层联邦学习应运而生[5]. 终端设备将本地更新的模型参数上传到边缘服务器用于中间聚合, 并上传至 SP 下的云服务器实现全局聚合.

在 HFL 中, 参与学习任务的终端设备需耗费计算、通信等资源. 如果没有足够的补偿, 终端设备无偿地贡献其资源是不切实际的[6], 同时, HFL 架构中存在梯度反演、搭便车攻击、成员推理攻击等安全隐患, 这些威胁可能导致终端设备数据部分泄露, 从而削弱终端设备的参与动机. 因此有必要设计一种有效的激励机制来激励终端设备参与并上传训练后的模型参数. 同时, 目前基于 HFL 的激励机制多采用博弈论的思想对参与方与联邦的关系进行建模, 来求解满足系统收益最大化的最优解. 然而, 通过整理相关研究可以发现, 现有工作假设终端设备、边缘服务器以及云服务器的博弈过程是信息对称的. 这种假设与 FL 保护终端设备隐私信息的初衷是相违背的[7].

针对上述问题, 在本文中提出了一种隐私保护的学习博弈 HFL 激励机制. 在边端层, 边缘服务器采用多维度契约论设计不同类型的多维合约鼓励终端设备在不泄露参与成本的前提下传输模型参数. 在云边层, 使用 Stackelberg 博弈建模边缘服务器与云服务器关于模型参数大小和其单位价格之间的交互关系, 并证明了存在唯一的斯塔克伯格均衡. 此外, 本文采用多智能体深度强化学习方法, 在保障边缘服务器和云服务器隐私的前提下逐步逼近 SE. 通过实验分析, 首先验证了提出的分层激励机制的有效性. 然后与其他基线方法比较, 提出的分层激励机制使云服务器的收益提升了接近 11%, 单位成本获取增益提升接近 18 倍.

2 相关工作

基于 HFL 激励机制设计[8,9]工作中, 文献[8]提出了 ODAM-DS 算法. 该算法基于在线双边拍卖机制, 优化边缘服务器在限定时间内选择移动设备, 以降低能耗. 针对实际场景中终端设备存在自私性的问题, 文献[9]提出一种基于博弈论的激励机制. 在激励预算有限的条件下, 得到了终端设备和边缘服务器之间的均衡解和最小的边缘模型训练时延. 然而上述文献聚焦于边缘服务器和终端设备之间关于激励策略的交互设计, 忽略了云服务中心和边缘服务器之间的策略交互设计.

在基于演化博弈激励机制设计[4,10]的现有工作中, 文献[10]提出了一个层次化博弈框架, 该框架利用演化博弈来模拟 FL 工作者与边缘服务器的动态关联, 并使用 Stackelberg 博弈来建模边缘服务器和模型所有者的最优带宽分配和奖励分配策略. 同时文献[4]针对 HFL 研发出资源分配及激励机制框架, 并应用演化博弈理论解决集群选择问题, 同时引入基于深度学习的拍卖机制评估集群头节点服务, 并通过性能评估证明了改方案的稳定性和效益.

在基于 Stackelberg 博弈激励机制设计[11–15]的现有工作中, 文献[11]将移动设备之间的交互构建为演化博弈, 同时将多边缘服务器之间的竞争构建为非合作博弈, 提出了基于多领导者 Stackelberg 博弈的激励机制, 该机制通过调整移动设备和边缘服务器的策略, 解决了收益的最优化问题. 文献[12]将端-边-云系统中的分层训练过程映射为通过实用函数相互连接的子博弈, 其中这些函数描述了每一层的实际价格情况, 并采用逆 Stackelberg 博弈理论方法来分析博弈过程, 进而指导参与者有效地获取所需的数据资源和利润. 文献[13]在边缘服务器和终端设备间使用联盟博弈进行边缘关联和带宽分配, 云中心和边缘服务器间采用 Stackelberg 博弈来优化边缘聚合次数和云服务器的奖励, 文献[14]

提出了三阶段 Stackelberg 博弈, 旨在模型所有者、集群头和无人机工作者之间建立一个层次化的互动模型, 以优化资源分配和学习效率. 文献[15]进一步基于 HFL 将单轮的激励机制优化为多轮的激励机制, 其中结合了基于拍卖的算法和组合多臂老虎机方法, 解决了客户端选择的挑战, 也最小化了 HFL 的训练延迟. 然而上述研究聚焦于对称信息下的激励机制设计, 其中对称信息的假设与 HFL 保护隐私的意图相违背. 因此, 在保护终端、边缘服务器和云服务器隐私的前提下, 本文提出了一种基于学习博弈和契约论的分层联邦学习激励机制.

3 系统模型

本节涉及的核心变量如表 1 所示.

表 1 核心变量注释表

符号	注释	符号	注释
M	边缘设备集合	$\lambda_{m,n}^{\mathrm{col}}$	终端设备n_m数据采集的单位成本
N	终端设备集合	$\lambda_{m,n}^{\mathrm{com}}$	终端设备n_m本地模型训练的能量消耗的单位成本
N_m	第m个边缘服务器内的终端集合	$\lambda_{m,n}^{\mathrm{upl}}$	终端设备n_m传输更新的本地模型参数的能量消耗的单位成本
n_m	第m个边缘服务器下第n个终端设备	$x_{m,n}$	终端设备n_m收集数据的比例
ξ_m	单位满意度所获得的利润	$\kappa_{m,n}$	有效电容开关
D	终端设备n_m可收集的最大数据量	$f_{m,n}$	表示终端设备n_m提供的算力
σ	终端设备n_m本地迭代次数	$d_{m,n}$	传输功率
s	单位数据量的比特数	$r_{m,n}$	终端设备n_m收到的来自第m个边缘服务器的奖励
η	终端设备n_m处理单位比特数的数据需要的算力	$\gamma_{m,i}$	类型$-(m,i,j,k)$终端设备数据采集成本类型
τ	边缘服务器m的迭代次数	$\nu_{m,j}$	类型$-(m,i,j,k)$终端设备模型训练成本类型
p	单位数据奖励	$\varphi_{m,k}$	类型$-(m,i,j,k)$终端设备模型传输成本类型
ψ	单位精度获得的利润	$t_{\max}$	本地模型训练以及更新后的模型参数上传至边缘服务器m的总时间
z	第z次全局聚合	$q_{m,i,j,k}$	边缘服务器m下某个终端设备属于类型$\gamma_{m,i}$, $\nu_{m,j}$, $\varphi_{m,k}$的概率
w^z	全局模型	$r_{m,i,j,k}$	类型$-(m,i,j,k)$终端设备收到的来自第m个边缘服务器的奖励
$w_{m,n}^{z,\sigma}$	终端设备n_m更新后的本地模型参数	$x_{m,i,j,k}$	类型$-(m,i,j,k)$终端设备收集数据的比例

3.1 系统模型

如图 1 所示, HFL 架构由一个云服务器, M个边缘设备, N个终端设备组成, 其中N_m表示第m个边缘服务器内的终端集合, n_m表示第m个边缘服务器下第n个终端设备, 考虑完全同步的 FL, 此架构完成一次全局训练过程如下.

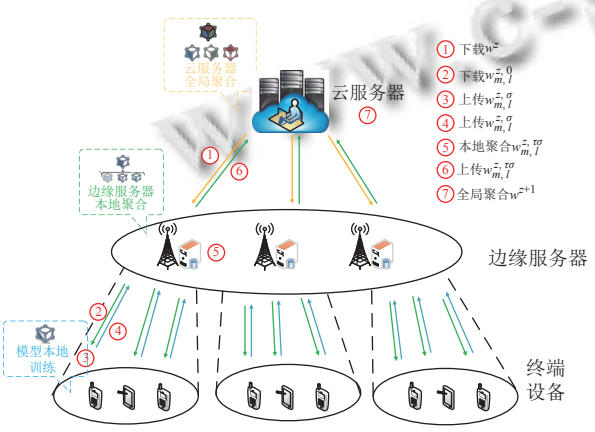

图 1 分层联邦学习架构图

(1) 终端设备: 终端设备n_m从边缘服务器接收全局模型w^z, 并利用本地数据$x_{m,n}D$进行模型训练. 其中z代表第z次全局聚合, D表示终端设备n_m可收集的最大数据量, $x_{m,n}$表示终端设备n_m收集数据的比例, 通过最小化损失函数训练完成后[5], 终端设备n_m将更新后的本地模型参数$w_{m,n}^{z,\sigma}$上传至边缘服务器m, 并获得来自边缘服务器m的奖励, 其中σ表示终端设备n_m本地迭代次数. 终端设备n_m在参与上述训练过程中会有一定成本, 即数据采集成本, 模型训练成本以及模型传输成本. 参考文献[16], 由于终端设备n_m需收集数据参与训练, 其数据采集成本定义如下:

$$C_{m,n}^{\mathrm{col}} = \lambda_{m,n}^{\mathrm{col}} x_{m,n} D \tag{1}$$

其中, $\lambda_{m,n}^{\mathrm{col}}$表示终端设备$n_m$数据采集的单位成本.

参考文献[16], 终端设备n_m本地模型训练成本定义如下:

$$C_{m,n}^{\mathrm{com}} = \lambda_{m,n}^{\mathrm{com}} D\sigma x_{m,n} s\eta\kappa_{m,n} f_{m,n}^2 \tag{2}$$

其中, $\lambda_{m,n}^{\mathrm{com}}$表示终端设备$n_m$本地模型训练的能量消耗的单位成本, s表示单位数据量的比特数, η表示终端设备n_m处理单位比特数的数据需要的算力, $\kappa_{m,n}$表示有效电容开关, $f_{m,n}$表示终端设备n_m提供的算力.

此外, 终端设备n_m会因与边缘服务器m通信产生一定的传输成本, 则模型传输成本定义如下:

$$C_{m,n}^{\mathrm{upl}}=\lambda_{m,n}^{\mathrm{upl}}d_{m,n}\left(t_{\max}-\frac{Dx_{m,n}\eta s}{f_{m,n}}\right) \tag{3}$$

其中, $\lambda_{m,n}^{\mathrm{upl}}$表示终端设备$n_m$传输更新的本地模型参数的能量消耗的单位成本, $d_{m,n}$表示传输功率, $t_{\max}$表示本地模型训练以及更新后的模型参数上传至边缘服务器m的总时间.

(2) 边缘服务器: 边缘服务器m对汇集的模型参数进行整合, 并将聚合后的参数$w_m^{z,\tau\sigma}=\sum_{n\in N_m}(x_{m,n}w_{m,n}^{z,\tau\sigma}/\sum_{n\in N_m}x_{m,n})$上传至云服务器, 其中$\tau$表示边缘服务器$m$的迭代次数, 作为回报, 边缘服务器$m$将从云服务器获得相应的奖励. 同时边缘服务器$m$利用契约论为终端设备$n_m$设计最优多维合约, 终端设备$n_m$选择使其收益最大化的多维合约.

(3) 云服务器: 在云端, 服务器对模型参数进行全局聚合, 得到新的全局模型为:

$$w^{z+1}=\sum_{m\in M}\frac{\sum_{n\in N_m}x_{m,n}}{\sum_{m\in M}\sum_{n\in N_m}x_{m,n}}w_m^{z,\tau\sigma} \tag{4}$$

一旦更新完成, 新的全局模型随即被发送至边缘服务器m, 边缘服务器进一步将其下发给终端设备n_m. 同时在确保模型性能满足的前提下, 云服务器会下发使其和边缘服务器m收益最大化的最优单位数据奖励. 由于终端设备n_m所提供的数据越多, 模型精度就越准确, 同时随着数据量的增加, 训练的全局模型的准确性会以较慢的速度提升, 参考文献[7,17], 本文将模型训练的准确性定义为:

$$A^{\mathrm{sp}}=\ln\left(1+\sum_{m\in M}\sum_{n\in N_m}Dx_{m,n}\right) \tag{5}$$

上述 3 个步骤将循环迭代, 直至全局模型收敛.

3.1.1 终端设备收益

边缘设备经过τ轮本地模型参数聚合后, 终端设备n_m会收到来自第m个边缘服务器的奖励$r_{m,n}$, 则终端设备n_m的收益定义如下:

$$\begin{aligned}u_{m,n}^{ed}=&r_{m,n}-\lambda_{m,n}^{\mathrm{col}}Dx_{m,n}-\lambda_{m,n}^{\mathrm{com}}\tau\sigma Dx_{m,n}s\eta\kappa_{m,n}f_{m,n}^2\\&-\lambda_{m,n}^{\mathrm{upl}}\tau d_{m,n}\left(t_{\max}-\frac{Dx_{m,n}\eta s}{f_{m,n}}\right)\end{aligned} \tag{6}$$

式 (6) 经过简化, 可重新定义如下:

$$u_{m,n}^{ed}=r_{m,n}-\gamma_{m,n}Dx_{m,n}-v_{m,n}Dx_{m,n}+\varphi_{m,n}Dx_{m,n}-\delta_{m,n} \tag{7}$$

其中, $\gamma_{m,n}=\lambda_{m,n}^{\mathrm{col}}$, $v_{m,n}=\lambda_{m,n}^{\mathrm{com}}\tau\sigma\eta s\kappa_{m,n}f_{m,n}^2$, $\varphi_{m,n}=\lambda_{m,n}^{\mathrm{upl}}\tau d_{m,n}\eta s/f_{m,n}$, $\delta_{m,n}=\tau\lambda_{m,n}^{3}t_{\max}d_{m,n}$, 为了简化分析, 参考文献[16], δ的值针对所有终端设备保持一致.

本文聚焦信息不对称场景. 边缘设备对于终端设备在参与 FL 任务的过程当中所产生的数据采集成本, 模型训练成本以及模型传输成本一概不知, 然而, 边缘设备可以通过历史记录获取到终端设备各个成本的分布情况, 因此边缘服务器m下某个终端设备可根据成本类型划分为 3 种, 分别定义为: $Y_m=\{\gamma_{m,i}:1\leqslant i\leqslant I_m\}$, $\Lambda_m=\{v_{m,j}:1\leqslant j\leqslant J_m\}$, $\Psi_m=\{\varphi_{m,k}:1\leqslant k\leqslant K_m\}$, 故第$m$个边缘服务器通信范围内共存在$I_mJ_mK_m$类型的终端设备. 边缘服务器$m$下某个终端设备属于类型$\gamma_{m,i}$, $v_{m,j}$, $\varphi_{m,k}$的概率定义为$q_{m,i,j,k}(\gamma_{m,i},v_{m,j},\varphi_{m,k})$, 可得出$\sum_{i\in I}\sum_{j\in J}\sum_{k\in K}N_mq_{m,i,j,k}=N_m$. 因此边缘服务器$m$下某个终端设备的成本类型可重新定义为$0<\gamma_{m,1}\leqslant\gamma_{m,2}\leqslant\cdots\leqslant\gamma_{m,I}$, $0<v_{m,1}\leqslant v_{m,2}\leqslant\cdots\leqslant v_{m,J}$, $0<\varphi_{m,1}\leqslant\varphi_{m,2}\leqslant\cdots\leqslant\varphi_{m,K}$, 边缘服务器$m$内上述 3 种成本类型的终端设备可定义为类型$-(m,i,j,k)$. 类型$-(m,i,j,k)$终端设备收益可重新定义如下:

$$\begin{aligned}u_{m,i,j,k}^{ed}=&r_{m,i,j,k}-\gamma_{m,i}Dx_{m,i,j,k}\\&-v_{m,j}Dx_{m,i,j,k}+\varphi_{m,k}Dx_{m,i,j,k}-\delta\end{aligned} \tag{8}$$

其中, $r_{m,i,j,k}$表示为类型$-(m,i,j,k)$终端设备收到的来自第m个边缘服务器的奖励, $x_{m,i,j,k}$表示为类型$-(m,i,j,k)$终端设备收集数据的比例.

3.1.2 边缘服务器收益

在 HFL 架构中, 云服务器给予边缘服务器的奖励越高, 边缘服务器的满意度也相应越高, 但随着奖励逐渐递增, 满意度的边际效用会逐渐下降. 因此边缘服务器m的满意度定义为$\sum_{i\in I}\sum_{j\in J}\sum_{k\in K}N_mq_{m,i,j,k}\ln(1+pDx_{m,i,j,k})$, 其中$p$表示单位数据奖励, 同时边缘服务器会下发奖励以此激励终端设备参与到 FL 任务中, 则边缘服务器的激励支出可定义为$\sum_{i\in I}\sum_{j\in J}\sum_{k\in K}N_mq_{m,i,j,k}r_{m,i,j,k}$, 最终边缘服务器$m$的收益定义如下:

$$\begin{aligned}u_m^{es}=&\sum_{i\in I}\sum_{j\in J}\sum_{k\in K}N_mq_{m,i,j,k}\xi_m\ln\left(1+pDx_{m,i,j,k}\right)\\&-\sum_{i\in I}\sum_{j\in J}\sum_{k\in K}N_mq_{m,i,j,k}r_{m,i,j,k}\end{aligned} \tag{9}$$

其中, ξ_m表示单位满意度所获得的利润.

3.1.3 云服务器收益

参考文献[7], 本文在A^{sp}的基础上, 将N个终端设备的平均增益定义为$\psi\ln(1+\sum_{m\in M}\sum_{i\in I}\sum_{j\in J}\sum_{k\in K}N_m q_{m,i,j,k}D x_{m,i,j,k})$, 其中$\psi$表示单位精度获得的利润. 云服务器负责分配激励给每个边缘服务器, 其激励成本定义为$\sum_{m\in M}\sum_{i\in I}\sum_{j\in J}\sum_{k\in K}pN_m q_{m,i,j,k}Dx_{m,i,j,k}$, 因此, 云服务器的收益定义如下:

$$u^{\mathrm{sp}}=\psi\ln\left(1+\sum_{m\in M}\sum_{i\in I}\sum_{j\in J}\sum_{k\in K}N_m q_{m,i,j,k}Dx_{m,i,j,k}\right)-\sum_{m\in M}\sum_{i\in I}\sum_{j\in J}\sum_{k\in K}pN_m q_{m,i,j,k}Dx_{m,i,j,k} \tag{10}$$

4 信息不对称下的激励机制

在 HFL 框架中, 终端设备面临着信息不完整的问题. 由于缺乏来自终端设备的先验信息, 云服务器无法准确判断哪些设备愿意参与模型训练, 其次边缘服务器对终端设备参与训练所产生的数据采集成本, 模型训练成本以及模型传输成本一概不知, 同时云服务器无法掌握终端设备本地模型训练的可用数据量, 因此无法给出给予边缘服务器最优的单位奖励. 最终以上问题会导致对终端设备提供激励时的过度支出.

为了解决上述问题, 本文基于 HFL 架构, 设计出一种分层激励机制. 在上层云服务器和边缘服务器间引入基于 MADRL 的 Stackelberg 博弈, 在下层边缘服务器和终端设备间引入契约论. 通过上述机制, 云服务器, 边缘服务器, 终端设备能够在各自信息不对称的前提下, 保证自身收益的最大化以及隐私的同时设计出最优的单位价格激励.

4.1 边端层契约论问题定义与求解

4.1.1 契约论问题定义

本文将 HFL 架构中下层终端设备与边缘服务器间的交互定义为契约论问题. 每个边缘服务器m为其交互范围内的所有终端设备根据其自身成本类型设计多维合约, 同时保证了自身收益的最大化. 其中合约包括边缘服务器提供给参与 FL 的终端设备的激励奖励r以及终端设备应当为模型训练提供的数据量x, 故将多维合约定义为$B(Y,A,\Psi)=\{\pi_{m,i,j,k}=\{r_{m,i,j,k},x_{m,i,j,k}\},1\leqslant i\leqslant I,1\leqslant j\leqslant J,1\leqslant k\leqslant K\}$.

为了保证合约的有效性与可行性, 即终端设备需确保选择符合自身成本类型的合约, 保证自身收益的最大化, 本文引入激励兼容条件 (incentive compatibility, IC), 同时终端设备需保证获取到自身期望的最低收益, 本文引入个体合理性条件 (individual rationality, IR)[18]. 对于类型$-(m,i,j,k)$的终端设备, 其 IR 定义如下:

$$u^{ed}_{m,i,j,k}\left(\pi_{m,i,j,k}\right)\geqslant 0,\ 1\leqslant i\leqslant I,1\leqslant j\leqslant J,1\leqslant k\leqslant K \tag{11}$$

IC 定义如下:

$$\begin{cases} u^{ed}_{m,i,j,k}\left(\pi_{m,i,j,k}\right)\geqslant u^{ed}_{m,i,j,k}\left(\pi_{m,i',j',k'}\right)\\ 1\leqslant i\leqslant I,1\leqslant j\leqslant J,1\leqslant k\leqslant K\\ 1\leqslant i'\leqslant I,1\leqslant j'\leqslant J,1\leqslant k'\leqslant K \end{cases} \tag{12}$$

在单位数据奖励固定的前提下, 每个边缘服务器作为合约的制定方需要在满足 IC 和 IR 条件的前提下最大化自己的收益, 故优化问题 P1 定义如下:

$$\begin{cases} \max_{r_{m,i,j,k},x_{m,i,j,k}} u^{es}_m,\ \forall m\in M\\ \text{s.t.}\begin{cases} \text{C1: 式 (11), 式 (12)}\\ \text{C2: } 0\leqslant r_{m,i,j,k}\leqslant 1,1\leqslant i\leqslant I,1\leqslant j\leqslant J,1\leqslant k\leqslant K\\ \text{C3: } 0\leqslant x_{m,i,j,k}\leqslant 1,1\leqslant i\leqslant I,1\leqslant j\leqslant J,1\leqslant k\leqslant K\\ 1\leqslant i'\leqslant I,1\leqslant j'\leqslant J,1\leqslant k'\leqslant K \end{cases} \end{cases} \tag{13}$$

4.1.2 契约论问题求解

优化问题 P1 涉及IJK个非凸 IR 约束和$IJK(IJK-1)$个非凸 IC 约束, 当涉及多个非凸条件时, 优化问题 P1 很难直接求解. 因此, 为了求解可行的多维合约, 需将多维合约先转化为单维合约. 基于式 (8), 类型为$-(m,i,j,k)$的终端设备的成本定义为$C_{m,i,j,k}(x_{m,i,j,k})=\gamma_{m,i}Dx_{m,i,j,k}+v_{m,j}Dx_{m,i,j,k}-\varphi_{m,k}Dx_{m,i,j,k}+\delta$, 进而推导出类型为$-(m,i,j,k)$的终端设备关于数据量的边际参与成本$\alpha$定义如下:

$$\alpha\left(\gamma_{m,i},v_{m,j},\varphi_{m,k}\right)=\frac{\partial C_{m,i,j,k}\left(x_{m,i,j,k}\right)}{\partial x_{m,i,j,k}}=D\gamma_{m,i}+Dv_{m,j}-D\varphi_{m,k} \tag{14}$$

其中, 当$\alpha\left(\gamma_{m,i},v_{m,j},\varphi_{m,k}\right)$非负且数值越大时, 表明具有较高边际参与成本的终端设备并不愿意参与到 HFL 任务当中.

为便于后续合约求解, 本文对$I_mJ_mK_m$个类型按照边际参与成本非递减顺序进行重新排序如下:

$$\Phi_{m,1}(x),\Phi_{m,2}(x),\cdots,\Phi_{m,h}(x),\cdots,\Phi_{m,IJK}(x) \tag{15}$$

其中, $\Phi_{m,h}(x)$表示某些类型$-(m,i,j,k)$终端设备属于$-(m,h)$类型的终端设备.

终端设备边际参与成本按照升序重新排列为:

$$\alpha(\Phi_{m,1},x)\leqslant\cdots\leqslant\alpha(\Phi_{m,h},x)\leqslant\cdots\leqslant\alpha(\Phi_{m,IJK},x)\quad(16)$$

为便于后续引理及定理的证明, 本文将类型为$\Phi_{m,h}(x)$的终端设备所对应的合约定义为$\pi_{m,h}=(x_{m,h},r_{m,h})$, 终端设备边际参与成本的排序重新定义为$C(\Phi_{m,h},x_{m,h})$.

下面将使用终端设备的边际参与成本类型来分析满足 IR 和 IC 条件的可行合约的充要条件.

本文参考文献[16], 得到如下的引理 1–引理 4.

引理 1. 对于任何可行的合约, 当且仅当$r_{m,i}<r_{m,j}$时, $x_{m,i}<x_{m,j},i,j\in[1,IJK]$.

引理 1 表明当合约可行时, 当终端设备贡献的数据量越多时所需要的奖励越多.

引理 2. 对于可行的合约$B(Y,\Lambda,\Psi)$, 如果对于任意x, $\alpha(\Phi_{m,i},x)>\alpha\left(\Phi_{m,j},x\right)$成立, 那么$x_{m,i}\leqslant x_{m,j}$.

引理 2 表明合约的单调性, 即当终端设备的边际参与成本越大时, 其越不愿意贡献数据.

由引理 1 和引理 2 可知, 可行合约的必要条件为:

$$\begin{cases}x_{m,1}\geqslant x_{m,2}\geqslant\cdots\geqslant x_{m,i}\geqslant\cdots\geqslant x_{m,IJK}\\ r_{m,1}\geqslant r_{m,2}\geqslant\cdots\geqslant r_{m,i}\geqslant\cdots\geqslant r_{m,IJK}\end{cases}\quad(17)$$

引理 3. 对于可行的合约, 如果类型为$\Phi_{m,IJK}$的终端设备的 IR 约束成立, 其余类型终端设备的 IR 约束随之成立.

引理 4. 对于可行的合约, 如果$\pi_{m,i-1}\overset{\text{PIC}}{\Leftrightarrow}\pi_{m,i}$并且$\pi_{m,i}\overset{\text{PIC}}{\Leftrightarrow}\pi_{m,i+1}$, 那么$\pi_{m,i-1}\overset{\text{PIC}}{\Leftrightarrow}\pi_{m,i+1}$.

通过引理 4 可将$IJK(IJK-1)/2$个 PIC 约束简化为$IJK-1$个 PIC 约束. 同时由引理 4 的证明可得合约可行的充分条件为:

$$\begin{cases}r_{m,IJK}-C(\Phi_{m,IJK},x_{m,IJK})\geqslant 0\\ r_{m,h+1}-C(\Phi_{m,h+1},x_{m,h+1})+C(\Phi_{m,h+1},x_{m,h})\geqslant\\ r_{m,h}\geqslant r_{m,h+1}-C(\Phi_{m,h},x_{m,h+1})+C(\Phi_{m,h},x_{m,h})\end{cases}\quad(18)$$

综上所述, 可得合约可行的充要条件如下.

定理 1. 在信息不对称场景下, 一个可行的合约必须满足如下 3 个条件.

$$\begin{cases}(1)\ \alpha(\Phi_{m,1},x)\leqslant\cdots\leqslant\alpha(\Phi_{m,h},x)\leqslant\cdots\leqslant\alpha(\Phi_{m,IJK},x),\\ \quad r_{m,IJK}\leqslant\cdots\leqslant r_1\leqslant 0,x_{m,IJK}\leqslant\cdots\leqslant x_1\leqslant 0\\ (2)\ r_{m,IJK}-C(\Phi_{m,IJK},x_{m,IJK})\geqslant 0\\ (3)\ r_{m,h+1}-C(\Phi_{m,h+1},x_{m,h+1})+C(\Phi_{m,h+1},x_{m,h})\\ \quad\geqslant r_{m,h}\geqslant r_{m,h+1}-C(\Phi_{m,h},x_{m,h+1})+C(\Phi_{m,h},x_{m,h})\end{cases}\quad(19)$$

证明: 基于引理 1 和 2 可得式 (19) 中的条件 (1), 引理 3 的数学表示形式即式 (19) 中的条件 (2), 基于引理 4 可得到式 (19) 中的条件 (3).

定理 2. 对于任意终端设备所提供用于训练的数据量, 可通过式 (20) 得到其最优奖励.

$$r_{m,h}^{*}=\begin{cases}C(\Phi_{m,h},x_{m,h}),\ h=IJK\\ r_{m,h+1}-C(\Phi_{m,h},x_{m,h+1})+C(\Phi_{m,h},x_{m,h}),\\ \quad h=1,\cdots,IJK-1\end{cases}\quad(20)$$

为便于求解合约, 参考文献[19,20], 定理 1 中的条件 (1) 和条件 (3) 需转化如下:

$$\begin{cases}(1)\ r_{m,IJK}-C(\Phi_{m,IJK},x_{m,IJK})=0\\ (2)\ r_{m,h+1}-C(\Phi_{m,h},x_{m,h+1})=r_{m,h}-C(\Phi_{m,h},x_{m,h}),\\ \quad h\in\{1,\cdots,IJK-1\}\end{cases}\quad(21)$$

令$\Delta_h=r_{m,h+1}-r_{m,h}=C(\varphi_{m,h},x_{m,h+1})-C(\varphi_{m,h},x_{m,h})$, $h=1,\cdots,IJK-1$以及$\Delta_{IJK}=0$. 则根据转化后的式 (21), 可将$r_{m,h}^{*}$改写为以下形式:

$$r_{m,h}^{*}=\begin{cases}C(\phi_{m,IJK},x_{m,IJK}),\quad h=IJK\\ C(\phi_{m,IJK},x_{m,IJK})-\sum_{k=h}^{IJK}\Delta_k,\ h=1,\cdots,IJK-1\end{cases}\quad(22)$$

为了得到终端设备用于模型训练的最优数据量, 需将式 (22) 代入优化问题 P1, 将其改写为新的优化问题 P2:

$$\begin{cases}\max_{x_{m,h}}u_m^{es}\\ \text{s.t. C1: } 0\leqslant x_{m,IJK}\leqslant\cdots\leqslant x_{m,h}\leqslant\cdots\leqslant x_{m,1}\leqslant 1\end{cases}\quad(23)$$

其中, $u_m^{es}=\sum_{h=1}^{l/K}N_m q_{m,h}\zeta_m\ln(1+pDx_{m,h})-\sum_{h=1}^{l/K}Dx_{m,h}$ $b_{m,h}$, 且当$h=1$时, $b_{m,h}=N_m q_{m,1}\alpha(\phi_{m,1},x_{m,1})$, 当$h=2,\cdots,IJK$时, $b_{m,h}=\alpha(\phi_{m,h},x_{m,h})\sum_{j=1}^{h}N_m q_{m,j}-\alpha(\phi_{m,h-1},x_{m,h})\sum_{j=1}^{h-1}N_m q_{m,j}$. 因此优化问题 P2 可视为凸优化问题, 并可根据优化问题 P2 中函数u_m对于$x_{m,h}$的一阶导数等于零求得最优数据量的解为:

$$x_{m,h}^{*}=\frac{N_m q_{m,h}\zeta_m}{Db_{m,h}}-\frac{1}{Dp}\quad(24)$$

最终将式 (24) 代入式 (22) 得到最优合约为$(r_{m,h}^{*},x_{m,h}^{*})$.

4.2 云边层基于学习的 Stackelberg 博弈问题定义与求解

4.2.1 基于学习的 Stackelberg 博弈问题定义

本文首先将 HFL 架构中上层云服务器与边缘服务器间的交互定义为 Stackelberg 博弈问题. 该问题中, 云服务器担任领导者, 且拥有一个最优单位价格激励.

边缘服务器作为追随者, 其根据云服务器所提供的最优单位价格激励, 决定终端设备参与模型训练所需要贡献的数据量[21,22]. 在信息对称的前提下, 二者不断调整策略, 以达到双方收益的最大化, 并证得存在唯一的 SE, 然而信息对称的假设与 HFL 架构保护隐私的初衷相违背, 故本文将 Stackelberg 博弈问题转化为马尔可夫决策过程 (Markov decision process, MDP), 并采用一种分布式且保护隐私的 MADRL 方法在信息不对称的情况下逐步逼近 SE.

定义 1. 对于任意一个边缘服务器$m \in M$和云服务器之间, 当且仅当满足以下条件时, 存在一组特定的最优解(X^*, p^*), 使得边缘服务器和云服务器之间存在唯一 SE.

$$\begin{cases} u^{cs}(x^*, p^*) \geqslant u^{cs}(x^*, p) \\ u_m^{es}(x_m^*, r_m^*, p^*) \geqslant u_m^{es}(x_m, r_m^*, p^*), \forall m \in M \end{cases} \tag{25}$$

其中, $X^* = [x_1^*, \cdots, x_m^*, \cdots, x_M^*]$, $x_m^* = [x_{m,1}^*, \cdots, x_{m,h}^*, \cdots, x_{m,IIK}^*]$, $r_m^* = [r_{m,1}^*, \cdots, r_{m,h}^*, \cdots, r_{m,IJK}^*]$.

为保证信息对称下唯一的 SE, 即云服务器能够选择最大化其收益的最优单位价格激励, 优化问题 P3 定义如下:

$$\begin{cases} \max_p u^{cs} \\ \text{s.t. } \max_{r_{m,i,j,k}, x_{m,i,j,k}} u_m^{es} \\ \forall m \in M, \forall i \in I, \forall j \in J, \forall k \in K \end{cases} \tag{26}$$

4.2.2 基于学习的 Stackelberg 博弈问题求解

针对优化问题 P3, 可得定理 3.

定理 3. 当 $2p(D + \sum_{m\in M}\sum_{h\in IJK}(N_{m,h}q_{m,h}^2\xi_m / b_{m,h})) > \sum_{m\in M}\sum_{h\in IJK} N_{m,h}q_{m,h}$ 时, SE 的解析解如下所示:

$$p^* = \frac{B_2 + \sqrt{B_2^2 + 4B_1B_3}}{2B_1} \tag{27}$$

$$x_{m,h}^* = \frac{N_m q_{m,h}\zeta_m}{Db_{m,h}} - \frac{1}{Dp},\ h = 1, \cdots, IJK \tag{28}$$

$$r_{m,h}^* = \begin{cases} C\left(\Phi_{m,h}, x_{m,h}^*\right), & h = IJK \\ r_{m,h+1}^* - C\left(\Phi_{m,h}, x_{m,h+1}^*\right) + C\left(\Phi_{m,h}, x_{m,h}^*\right), & h = 1, \cdots, IJK - 1 \end{cases} \tag{29}$$

其中, $B_1 = D + \sum_{m\in M}\sum_{h\in IJK}(N_{m,h}q_{m,h}^2\xi_m / b_{m,h})$, $B_2 = \sum_{m\in M}\sum_{h\in IJK} N_{m,h}q_{m,h}$, $B_3 = \psi\sum_{m\in M}\sum_{h\in IJK} N_{m,h}q_{m,h} / \sum_{m\in M}\sum_{h\in IJK}(N_{m,h}q_{m,h}^2\xi_m / Db_{m,h})$.

证明: 见附录 A.

4.2.2.1 信息非对称下基于学习的智能决策方法

在本节中, 将探讨如何在信息不对称下逐渐逼近 SE. 由定理 3 可知, 在信息对称的下云服务器和边缘服务器之间存在唯一的 SE, 然而该分析是基于集中决策的, 即需要收集所有边缘服务器的隐私信息来确定 SE 的存在, 这与 HFL 框架保护隐私的初衷相违背. 为了解决上述问题, 参考文献[23–26], 本文将 Stackelberg 博弈问题转化为 MDP, 并采用 MADRL 方法在信息不对称的情况下逐步逼近 SE.

为了在隐私保护的情况下逼近 SE, 云服务器与 M 个边缘服务器都会成为 MADRL 中的一个智能体, 每个智能体需要基于不完全信息做出决策. 这些智能体组成的集合为$M = [0, 1, 2, \cdots, M]$, 其中智能体$m = 0$代表云服务器, 其他智能体对应不同的边缘服务器. 一个智能体的训练过程被划分为T次迭代, 在每次迭代t中, 云服务器以博弈的领导者身份与环境交互, 通过获取的状态S_0^t, 确定动作a_0^t, 而边缘服务器m以博弈从方身份通过与环境交互获取状态S_m^t, 确定动作a_m^t, 并通过收集所有参与方的策略分别计算它们的奖励G_0^t和G_m^t, 之后根据收集的策略来创建下一时刻的状态. 同时, 智能体会将所有操作的历史数据保存到一个大小为L的缓冲队列中, d个时隙之后, 将缓冲队列中的经验回放计算奖励, 以更新网络参数. 由于在每次迭代训练中, 智能体的状态是根据动作决策生成的, 并没有获取任何隐私信息, 这样就可以在没有隐私泄露的情况下逼近 SE.

4.2.2.2 马尔可夫决策过程

状态空间: 在 Stackelberg 博弈中, 每个智能体的状态是参考其他智能体的历史策略来确定的. 在第t次训练步骤中, 每个边缘服务器m (即智能体$m > 0$) 与环境互动, 以获得历史信息并了解当前状态, 此状态定义为$S_m^t = \{x_{-m}^{t-L}, p^{t-L}, x_{-m}^{t-L+1}, p^{t-L+1}, \cdots, x_{-m}^{t-1}, p^{t-1}\}$, 其中$x_{-m} = [x_1, \cdots, x_{m-1}, x_{m+1}, \cdots, x_M]$、$L$代表历史记录的长度、$x_m = [x_{m,1}, \cdots, x_{m,h}, \cdots, x_{m,IJK}]$, 对于云服务器, 其状态包括所有边缘服务器在过去L次历史动作, 故可将此状态定义为$S_0^t = \{x^{t-L}, x^{t-L+1}, \cdots, x^{t-1}\}$.

动作空间: 由于智能体的行动由决策变量所定义, 故在第t次训练步骤中, 云服务器的行动直接与决策变量相关, 可表示为$a_0^t = p$. 为加速智能体的学习过程, 需

为行动设置一个上限$p^{\max}$, 其中$p^t \in [0, p^{\max}]$. 对于每个边缘服务器, 其动作空间可表示为$a_m^t = [x_{m,1}^t, x_{m,h}^t, \cdots, x_{m,IIK}^t]$. 由于每个智能体的行动空间是连续的, 可将每个行动除以其最大值, 使云服务器和边缘服务器所对应的两个行动空间统一映射到区间[0,1].

奖励函数: 当所有策略执行完毕后, 每个智能体通过观察其他玩家的行为, 并根据收益函数来决定自身的即时奖励, 此奖励定义如下:

$$G_n^t = \begin{cases} u^{cs,t}, & m = 0 \\ u_m^{es,t}, & m \geqslant 1 \end{cases} \tag{30}$$

其中, 在第t次训练步骤中, $u_m^{es,t}$代表优化问题 P3 中每个边缘服务器的收益. 此时环境会更新每个代理的新状态, 并进入下一次训练$t+1$.

4.2.2.3 actor-critic 网络与策略优化

本文在每个智能体上设计了一个 actor 网络π_θ和 critic 网络v_σ, 通过学习π_θ来近似策略函数, 学习v_σ来近似值函数, 其中θ和σ是 actor 网络和 critic 网络的参数. 更进一步, 对于智能体m使用$V(S_m^t;\pi_{\theta_m})$表示状态价值函数, 使用$Q(S_m^t, a_m^t;\pi_{\theta_m})$表示动作价值函数, 将智能体$m$的学习目标记为$\mathcal{L}_m$, 那么可以将它的学习目标表述为:

$$\begin{aligned} \theta_0^* &= \text{argmax}_{\theta_0} \mathcal{L}_0(\pi_{\theta_0}) = \text{argmax}_{\theta_0} \mathbb{E}\left[V(s_0^t;\pi_{\theta_0})\right] \\ &= \text{argmax}_{\theta_0} \mathbb{E}\left[Q(s_0^t, a_0^t;\pi_{\theta_0})\right] \end{aligned} \tag{31}$$

本文中采用文献[23]中的近端策略优化裁剪算法来执行训练过程. 具体来说策略梯度被定义为:

$$\begin{aligned} \theta_m^* &= \text{argmax}_{\theta_m} \mathcal{L}_m(\pi_{\theta_m}) \\ &= \text{argmax}_{\theta_m} \mathbb{E}\left[V(S_m^t;\pi_{\theta_m})\right] \\ &= \text{argmax}_{\theta_m} \mathbb{E}\left[Q(S_m^t, a_m^t;\pi_{\theta_m})\right] \end{aligned} \tag{32}$$

$$C_{\pi_\theta}^m(S_m, a_m) = \min[P^m A_{\pi_\theta}^m(S_m, a_m), \mathcal{F}(P^m) A_{\pi_\theta}^m(S_m, a_m)] \tag{33}$$

$$P^m = \frac{\pi_\theta^m(S_m \mid a_m)}{\hat{\pi}_\theta^m(S_m \mid a_m)} \tag{34}$$

$$A_{\pi_\theta}^m(S_m, a_m) = Q_{\pi_\theta}^m(S_m \mid a_m) - V_{\pi_\theta}^m(S_m) \tag{35}$$

$$\mathcal{F}(P^m) = \begin{cases} 1+\varepsilon, & P^m > 1+\varepsilon \\ P^m, & 1-\varepsilon < P^m < 1+\varepsilon \\ 1-\varepsilon, & P^m < 1-\varepsilon \end{cases} \tag{36}$$

ε是可调参数, 之后分别使用随机梯度上升和梯度下降更新 actor-critic 网络模型, 随着训练过程的继续, 智能体逐渐学习到最优策略. 当训练过程收敛时, 智能体根据行动者网络的输出来确定策略.

5 仿真实验

为评估本文提出的激励机制, 实验设计 1 个云服务器和 3 个边缘服务器, 每个边缘服务器连接 8 个终端设备. 终端设备的每种类型的取值范围是[1, 5], 同时每种类型的数量设置为$I = J = K = 2$. 每种类型的概率分布服从均匀分布, 即$q = 1/8$. 参考文献[27–30], 详细的配置参数见表 2.

表 2 仿真参数设置

参数	设置
图片大小	$s = 32\times32\times3\times8$ bit
有效开关电容	$\kappa = 10^{-28}$
计算1比特数据需要的CPU周期数	$\eta = 20$ cycles/bit
本地迭代次数	$\tau = 5$
传输功率	$p = 0.2$ W
最大时延	$t_{\max} = 0.01$ s
CPU周期数	$f = 10^9$ Hz
其他参数	$\beta^1 = 1, \beta^2 = 5\times10^3, \beta^3 = 10^2$

5.1 激励机制有效性验证

图 2 展示了不同单位利润下单位价格和云服务器收益之间的关系. 当单位利润固定时, 随着单位价格的增加, 云服务器收益先增加后减少. 即云服务器的收益关于单位价格是严格凹函数的, 存在最大值. 间接验证了定理 3 的有效性, 因为当云服务器和边缘服务器逐渐逼近 SE 时, 云服务器的收益也会逐渐达到最大值, 同时验证了激励机制的有效性.

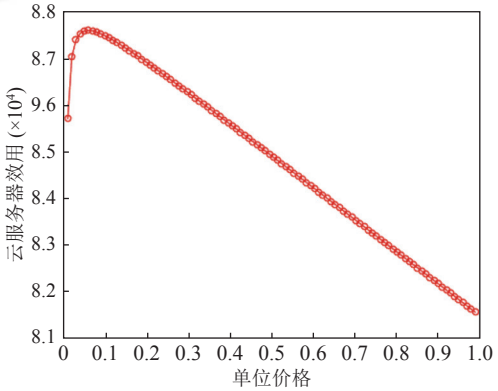

图 2 云服务器收益与单位价格的关系

图 3–图 5 展示了云服务器, 边缘服务器 1, 边缘服务器 2 的策略收敛情况. 其中云服务器的策略是单位价格, 而边缘服务器 1 和 2 的策略则基于终端设备提

供的数据量. 随着迭代次数的增加, 云服务器、边缘服务器和终端设备的策略不断演化, 直至收敛. 结果显示, 所有参与者的策略最终收敛于理论 SE 值, 且这一收敛状态在约 600 次训练集后达成. 这一结果表明, MADRL 方法极大地促进了参与者在合理时间内学习并找到近似最优策略.

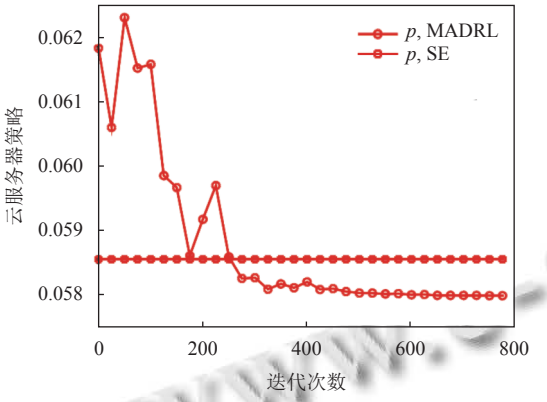

图 3 迭代次数与云服务器策略的关系

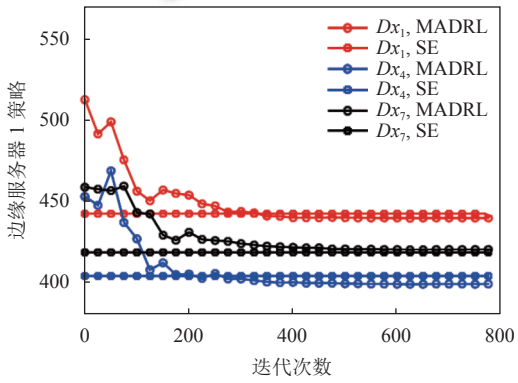

图 4 迭代次数与边缘服务器 1 策略的关系

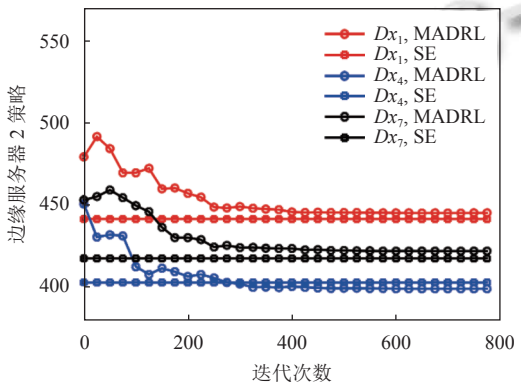

图 5 迭代次数与边缘服务器 2 策略的关系

图 6–图 7 展示了在信息不对称 (IR 和 IC) 的情况下, 本文所设计的多维合约针对成本类型为 1、3、5、7 的终端设备的激励有效性的验证. 即当终端设备选择适合其类型的合约时, 自身的收益都会达到最大, 例如当类型 3 的终端设备选择专为其设计的合约条款 (x_3, r_3) 时, 其收益最高, 如果选择其他合约条款, 其收益会降低, 这表明满足 IC 约束. 同时每个终端设备都能保证获取到自身期望最低收益, 即收益非负, 这表明满足 IR 约束.

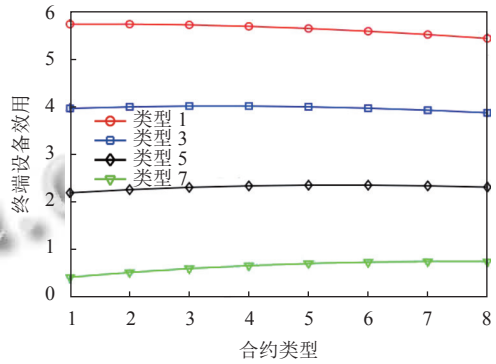

图 6 终端设备收益与合约类型的关系

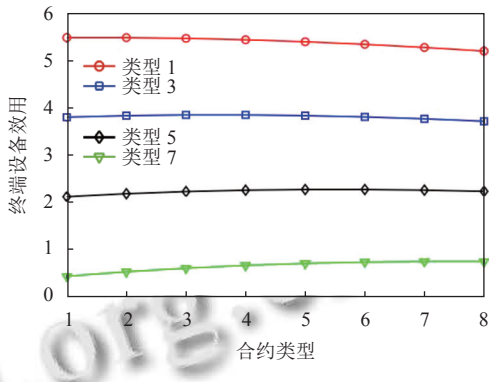

图 7 终端设备收益与合约类型的关系

5.2 激励机制对比

现有的工作设计了基于双层斯坦克尔伯格博弈 (Stackelberg-Stackelberg, SS) 激励机制[7], 其中终端设备收益定义如下:

$$u_{m,i,j,k}^{ed,SG}=\frac{Dx_{m,i,j,k}^{SG}R_m^{SG}}{\sum_{i\in I}\sum_{j\in J}\sum_{k\in K}Dx_{m,i,j,k}^{SG}}-\gamma_{m,l}Dx_{m,i,j,k}^{SG}-v_{m,j}Dx_{m,l,j,k}^{SG}+\varphi_{m,k}Dx_{m,i,j,k}^{SG} \tag{37}$$

边缘服务器收益定义如下:

$$u_m^{es,SG}=\sum_{i\in I}\sum_{j\in J}\sum_{k\in K}N_m q_{m,i,j,k}\xi_m\ln\left(1+p^{SG}Dx_{m,i,j,k}^{SG}\right)-\mathrm{R}_m^{SG} \tag{38}$$

云服务器的收益定义如下:

$$u^{cs,SG}=\psi\ln\left(1+\sum_{m\in M}\sum_{i\in I}\sum_{j\in J}\sum_{k\in K}N_m q_{m,i,j,k}Dx^{SG}_{m,i,j,k}\right)-\sum_{m\in M}\sum_{i\in I}\sum_{j\in J}\sum_{k\in K}p^{SG}N_m q_{m,i,j,k}Dx^{SG}_{m,i,j,k} \tag{39}$$

本文将提出的基于学习博弈-契约 (learning-based Stackelberg-contract, LSC) 激励机制与 SS 激励机制进行比较.

从图 8 可以看出, 与现有的 SS 激励机制方案, 本文提出的 LSC 激励机制方案能够使云服务器从边缘服务器和终端设备获得更多的收益, 收益提升了接近 11%. 同时边缘服务器和终端设备获得的收益是较低的, 原因在于 SS 激励机制方案考虑了边缘服务器与终端设备收益的最大化, 因而降低了云服务器的收益. 参考文献[7], 其实验结果显示云中心获取的收益与分层联邦学习的模型精度成正相关. 因此本文提出的 LSC 激励机制方案更有助于提升 HFL 模型精度.

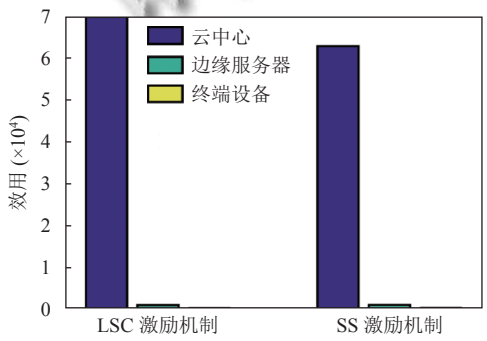

图 8 不同激励机制方法下云-边-端的收益对比

由于本文 LSC 激励机制方案与已有 SS 激励机制方案增益和总成本支出各不相同, 故为了进一步展示本文 LSC 激励机制方案的优越性, 本文在相同成本支出的情况下, 进一步比较两种方案获取的增益, 并将试验指标定义为增益除以总成本. 图 9 展示出本文提出的 LSC 激励机制方案能够使云服务器从边缘服务器和终端设备获得更多的单位成本获取增益, 提升接近 18 倍. 同时由式 (10) 可知, 增益和数据量成正比, 因此 LSC 激励机制方案在相同成本支出情况下, 能使云服务器获取更多的数据量, 从而分层联邦学习中云服务器的模型精度能够提升更高.

5.3 参数对性能的影响

图 10–图 12 探讨了单元利润和用户类型的变化对系统性能的影响. 结果显示, 提升单元利润显著增加了云服务器及边缘服务器与终端设备的整体收益. 同时, 用户类型的增加也相应提高了云服务器及边缘设备的整体收益.

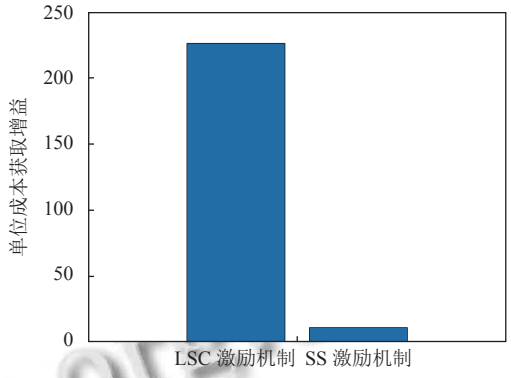

图 9 不同激励机制方法下云-边-端的单位成本获取增益对比

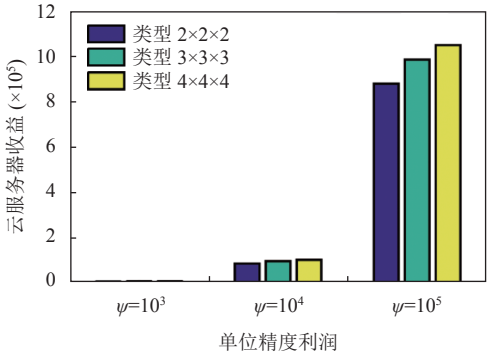

图 10 类型数量和单位精度利润对云服务器收益的影响

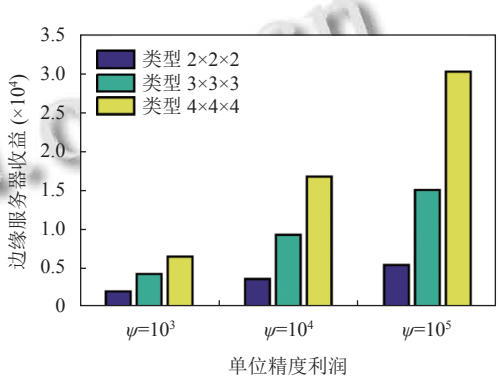

图 11 类型数量和单位精度利润对边缘服务器收益的影响

6 结论与展望

针对隐私保护下的 HFL 资源分配问题, 本文设计了一种保护参与者隐私的分层激励机制. 在边缘-设备层, 边缘服务器作为中间商利用多维契约论设计不同的合约激励不同类型的终端设备参与到 HFL 任务当中. 在服务提供商-边缘层中, 本文使用 Stackelberg 博弈来说明服务提供商下云服务器和边缘服务器之间的

交互, 证明在信息对称的情况下, 边缘服务器和云服务器之间存在唯一 SE. 此外, 本文采用分布式 MADRL 方法, 在信息不对称的情况下, 迭代收敛到 SE. 实验表明本文所提出的激励机制可以取得比其他基准方案更好的性能, 使云服务器的收益提升接近 11%, 单位成本获取增益提升接近 18 倍. 本文探究了单任务下的 HFL 激励机制. 在未来的工作中会尝试从多任务的角度出发, 使用多维双层契约理论方法设计 HFL 激励机制以达到提高其效率的目的.

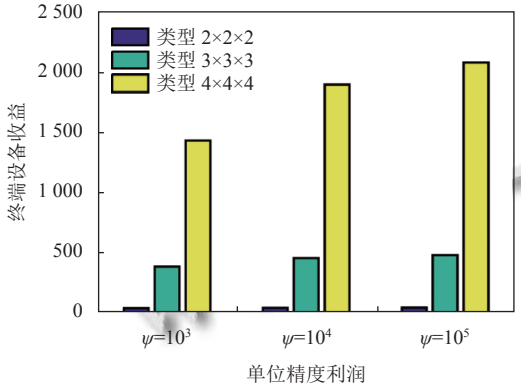

图 12 类型数量和单位精度利润对终端设备收益的影响

参考文献

1 Aledhari M, Razzak R, Parizi RM, *et al*. Federated learning: A survey on enabling technologies, protocols, and applications. IEEE Access, 2020, 8: 140699–140725. [doi: 10.1109/ACCESS.2020.3013541]

2 Lim WYB, Luong NC, Hoang DT, *et al*. Federated learning in mobile edge networks: A comprehensive survey. IEEE Communications Surveys & Tutorials, 2020, 22(3): 2031–2063.

3 Wei K, Li J, Ding M, *et al*. User-level privacy-preserving federated learning: Analysis and performance optimization. IEEE Transactions on Mobile Computing, 2022, 21(9): 3388–3401. [doi: 10.1109/TMC.2021.3056991]

4 Lim WYB, Ng JS, Xiong ZH, *et al*. Decentralized edge intelligence: A dynamic resource allocation framework for hierarchical federated learning. IEEE Transactions on Parallel and Distributed Systems, 2022, 33(3): 536–550. [doi: 10.1109/TPDS.2021.3096076]

5 Luo SQ, Chen X, Wu Q, *et al*. HFEL: Joint edge association and resource allocation for cost-efficient hierarchical federated edge learning. IEEE Transactions on Wireless Communications, 2020, 19(10): 6535–6548. [doi: 10.1109/TWC.2020.3003744]

6 Kang JW, Xiong ZH, Niyato D, *et al*. Incentive mechanism for reliable federated learning: A joint optimization approach to combining reputation and contract theory. IEEE Internet of Things Journal, 2019, 6(6): 10700–10714. [doi: 10.1109/JIOT.2019.2940820]

7 Wang XF, Zhao YF, Qiu C, *et al*. InFEDge: A blockchain-based incentive mechanism in hierarchical federated learning for end-edge-cloud communications. IEEE Journal on Selected Areas in Communications, 2022, 40(12): 3325–3342. [doi: 10.1109/JSAC.2022.3213323]

8 杜辉, 李卓, 陈昕. 基于在线双边拍卖的分层联邦学习激励机制. 计算机科学, 2022, 49(3): 23–30. [doi: 10.11896/jsjkx.210800051]

9 贾云健, 黄宇, 梁靓, 等. 基于主从博弈的分层联邦学习激励机制研究. 电子与信息学报, 2023, 45(4): 1366–1373. [doi: 10.11999/JEIT220175]

10 Lim WYB, Ng JS, Xiong ZH, *et al*. Dynamic edge association and resource allocation in self-organizing hierarchical federated learning networks. IEEE Journal on Selected Areas in Communications, 2021, 39(12): 3640–3653. [doi: 10.1109/JSAC.2021.3118401]

11 耿方兴, 李卓, 陈昕. 基于多领导者 Stackelberg 博弈的分层联邦学习激励机制设计. 计算机应用, 2023, 43(11): 3551–3558. [doi: 10.11772/j.issn.1001-9081.2022111727]

12 Zhao YF, Liu ZC, Qiu C, *et al*. An incentive mechanism for big data trading in end-edge-cloud hierarchical federated learning. Proceedings of the 2021 IEEE Global Communications Conference (GLOBECOM). Madrid: IEEE, 2021: 1–6.

13 Chu SF, Li J, Wei K, *et al*. Design of two-level incentive mechanisms for hierarchical federated learning. arXiv: 2304.04162, 2023.

14 He WJ, Yao HP, Mai T, *et al*. Three-stage stackelberg game enabled clustered federated learning in heterogeneous UAV swarms. IEEE Transactions on Vehicular Technology, 2023, 72(7): 9366–9380. [doi: 10.1109/TVT.2023.3246636]

15 Su LN, Li ZP. Incentive-driven long-term optimization for hierarchical federated learning. Computer Networks, 2023, 234: 109944. [doi: 10.1016/j.comnet.2023.109944]

16 Wu MQ, Ye DD, Ding JH, *et al*. Incentivizing differentially private federated learning: A multidimensional contract approach. IEEE Internet of Things Journal, 2021, 8(13): 10639–10651. [doi: 10.1109/JIOT.2021.3050163]

17 Ding NN, Fang ZX, Huang JW. Optimal contract design for efficient federated learning with multi-dimensional private information. IEEE Journal on Selected Areas in Communications, 2021, 39(1): 186–200. [doi: 10.1109/JSAC.2020.3036944]

18 Gao L, Wang XB, Xu YY, *et al*. Spectrum trading in cognitive radio networks: A contract-theoretic modeling approach. IEEE Journal on Selected Areas in Communications, 2011, 29(4): 843–855. [doi: 10.1109/JSAC.2011.110415]

19 Lim WYB, Huang JQ, Xiong ZH, *et al*. Towards federated learning in UAV-enabled internet of vehicles: A multi-dimensional contract-matching approach. IEEE Transactions on Intelligent Transportation Systems, 2021, 22(8): 5140–5154. [doi: 10.1109/TITS.2021.3056341]

20 Xiong ZH, Kang JW, Niyato D, *et al*. A multi-dimensional contract approach for data rewarding in mobile networks. IEEE Transactions on Wireless Communications, 2020, 19(9): 5779–5793. [doi: 10.1109/TWC.2020.2997023]

21 Maharjan S, Zhu QY, Zhang Y, *et al*. Dependable demand response management in the smart grid: A Stackelberg game approach. IEEE Transactions on Smart Grid, 2013, 4(1): 120–132. [doi: 10.1109/TSG.2012.2223766]

22 Başar T, Jan Olsder G. Dynamic noncooperative game theory. 2nd ed., Philadelphia: Society for Industrial and Applied Mathematics, 1999. 160.

23 Huang XM, Li PC, Yu R, *et al*. FedParking: A federated learning based parking space estimation with parked vehicle assisted edge computing. IEEE Transactions on Vehicular Technology, 2021, 70(9): 9355–9368. [doi: 10.1109/TVT.2021.3098170]

24 Zhan YF, Guo S, Li P, *et al*. A deep reinforcement learning based offloading game in edge computing. IEEE Transactions on Computers, 2020, 69(6): 883–893. [doi: 10.1109/TC.2020.2969148]

25 Huang XM, Zhong YP, Wu Y, *et al*. Privacy-preserving incentive mechanism for platoon assisted vehicular edge computing with deep reinforcement learning. China Communications, 2022, 19(7): 294–309. [doi: 10.23919/JCC.2022.07.022]

26 Li B, Xie K, Huang XM, *et al*. Deep reinforcement learning based incentive mechanism design for platoon autonomous driving with social effect. IEEE Transactions on Vehicular Technology, 2022, 71(7): 7719–7729. [doi: 10.1109/TVT.2022.3164656]

27 Zhou ZY, Liu PJ, Feng JH, *et al*. Computation resource allocation and task assignment optimization in vehicular fog computing: A contract-matching approach. IEEE Transactions on Vehicular Technology, 2019, 68(4): 3113–3125. [doi: 10.1109/TVT.2019.2894851]

28 Wang SM, Ye DD, Huang XM, *et al*. Consortium blockchain for secure resource sharing in vehicular edge computing: A contract-based approach. IEEE Transactions on Network Science and Engineering, 2021, 8(2): 1189–1201. [doi: 10.1109/TNSE.2020.3004475]

29 Li PC, Huang XM, Pan M, *et al*. FedGreen: Federated learning with fine-grained gradient compression for green mobile edge computing. Proceedings of the 2021 IEEE Global Communications Conference (GLOBECOM). Madrid: IEEE, 2021: 1–6.

30 Hou ZW, Chen H, Li YH, *et al*. Incentive mechanism design for wireless energy harvesting-based Internet of Things. IEEE Internet of Things Journal, 2018, 5(4): 2620–2632. [doi: 10.1109/JIOT.2017.2786705]

附录 A

优化问题 3 中, 关于收益的函数u_m的一阶偏导数为:

$$\frac{\partial u_m^{es}}{\partial x_{m,h}}=\frac{DN_m q_{m,h} p\zeta_m}{1+pDx_{x,h}}-b_{m,h} \tag{A1}$$

二阶偏导数为:

$$\frac{\partial u_m^{es,2}}{\partial x_{m,h}^2}=-\frac{N_m q_{m,h}\zeta_m D^2 p^2}{(1+pDx_{x,h})^2}<0 \tag{A2}$$

由式 (A2) 可得, 边缘服务器m的收益函数是一个严格的凹函数. 可令式 (A1) 等于0, 得到终端设备用于本地模型训练的最优数据量为$x_{m,h}^*=N_{mm}q_{m,h}\zeta_m/Db_{m,h}-1/Dp$, 然后将式 (24) 代入到式 (20) 可得终端设备最优奖励为$r_{m,h}^*$. 为证明信息对称下存在唯一 SE, 需将所求到的最优数据量$x_{m,h}^*$代入式 (9) 可得到单位价格奖励p关于收益的一个函数:

$$u^{es}=\psi\ln\left(1+\sum_{m\in M}\sum_{h\in IJK}N_m q_{m,h}\left(\frac{N_m q_{m,h}\zeta_m}{Db_{m,h}}-\frac{1}{Dp}\right)\right)-\sum_{m\in M}\sum_{h\in IJK}pN_m q_{m,h}\left(\frac{N_m q_{m,h}\zeta_m}{Db_{m,h}}-\frac{1}{Dp}\right) \tag{A3}$$

式 (A3) 的一阶偏导数为:

$$\begin{aligned}\frac{\delta u^{es}}{\delta p}&=\frac{\psi\sum_{m\in M}\sum_{h\in IJK}N_{m,h}q_{m,h}}{Dp^2\left[1+\sum_{m\in M}\sum_{h\in IJK}N_{m,h}q_{m,h}\left(\frac{q_{m,h}\xi_m}{Db_{m,h}}-\frac{1}{Dp}\right)\right]}-\sum_{m\in M}\sum_{h\in IJK}\frac{N_{m,h}q_{m,h}^2\xi_m}{Db_{m,h}}\\&=\frac{\psi\sum_{m\in M}\sum_{h\in IJK}N_{m,h}q_{m,h}}{p^2\left(D+\sum_{m\in M}\sum_{h\in IJK}\frac{N_{m,h}q_{m,h}^2\xi_m}{b_{m,h}}\right)-p\sum_{m\in M}\sum_{h\in IJK}N_{m,h}q_{m,h}}-\sum_{m\in M}\sum_{h\in IJK}\frac{N_{m,h}q_{m,h}^2\xi_m}{Db_{m,h}}\end{aligned}\tag{A4}$$

式 (A3) 的二阶偏导数为:

$$\begin{aligned}\frac{\delta^2 u^{es}}{\delta p^2}=&-\frac{\psi\sum_{m\in M}\sum_{h\in IJK}N_{m,h}q_{m,h}\left(2p\left(D+\sum_{m\in M}\sum_{h\in IJK}\frac{N_{m,h}q_{m,h}^2\xi_m}{b_{m,h}}\right)\right)}{\left[p^2\left(D+\sum_{m\in M}\sum_{h\in IJK}\frac{N_{m,h}q_{m,h}^2\xi_m}{b_{m,h}}\right)-p\sum_{m\in M}\sum_{h\in IJK}N_{m,h}q_{m,h}\right]^2}\\&+\frac{\psi\sum_{m\in M}\sum_{h\in IJK}N_{m,h}q_{m,h}\left(\sum_{m\in M}\sum_{h\in IJK}N_{m,h}q_{m,h}\right)}{\left[p^2\left(D+\sum_{m\in M}\sum_{h\in IJK}\frac{N_{m,h}q_{m,h}^2\xi_m}{b_{m,h}}\right)-p\sum_{m\in M}\sum_{h\in IJK}N_{m,h}q_{m,h}\right]^2}\end{aligned}\tag{A5}$$

当 $2p(D+\sum_{m\in M}\sum_{h\in IJK}N_{m,h}q_{m,h}^2\xi_m/b_{m,h})>\sum_{m\in M}\sum_{h\in IJK}N_{m,h}q_{m,h}$时, 式 (A5)<0, 则可求得最优单位价格激励, 也因此存在唯一 SE. 当且仅当 $B_1=D+\sum_{m\in M}\sum_{h\in IJK}N_{m,h}q_{m,h}^2\xi_m/b_{m,h}$, $B_2=\sum_{m\in M}\sum_{h\in IJK}N_{m,h}q_{m,h}$ 和 $B_3=\psi\sum_{m\in M}\sum_{h\in IJK}N_{m,h}q_{m,h}/\sum_{m\in M}\sum_{h\in IJK}(N_{m,h}q_{m,h}^2\xi m/Db_{m,h})$ 时, 式 (A4)=0, 可求得 $p_1=B_2+\sqrt{B_2^2+4B_1B_3}/2B_1$, $p_2=B_2-\sqrt{B_2^2+4B_1B_3}/2B_1$, 而当 $B_2<\sqrt{B_2^2+4B_1B_3}$时, $p_2<0$. 因此, 云服务器所下发的最优单位价格激励为 $p^*=p_1=B_2+\sqrt{B_2^2+4B_1B_3}/2B_1$.

(校对责编: 孙君艳)