智能系统学报
CAAI TRANSACTIONS ON INTELLIGENT SYSTEMS

基于数据质量评估的高效强化联邦学习节点动态采样优化

赵泽华, 梁美玉, 薛哲, 李昂, 张珉

引用本文:
赵泽华, 梁美玉, 薛哲, 等. 基于数据质量评估的高效强化联邦学习节点动态采样优化[J]. 智能系统学报, 2024, 19(6): 1552-1561.
ZHAO Zehua, LIANG Meiyu, XUE Zhe, et al. Client dynamic sampling optimization of efficient reinforcement federated learning based on data quality assessment[J]. *CAAI Transactions on Intelligent Systems*, 2024, 19(6): 1552-1561.

在线阅读 View online: https://dx.doi.org/10.11992/tis.202305054

DOI: 10.11992/tis.202305054
网络出版地址: https://link.cnki.net/urlid/23.1538.TP.20240912.1131.004

基于数据质量评估的高效强化联邦学习节点动态采样优化

赵泽华, 梁美玉, 薛哲, 李昂, 张珉
(北京邮电大学 智能通信软件与多媒体北京市重点实验室 北京 100876)

摘 要: 系统异构性和统计异构性的存在使得通信开销和通信效率成为联邦学习的关键瓶颈之一, 在众多参与方中只选取一部分客户端执行模型更新和聚合可以有效降低通信开销, 但是选择偏差和客户端上的数据质量分布不平衡对客户端采样方法提出了额外的挑战。为此, 提出数据质量评估的高效强化联邦学习节点动态采样优化方法(client dynamic sampling optimization of efficient reinforcement federated learning based on data quality assessment, RQCS), 该方法采用沙普利值的贡献指数评估客户端上的数据质量, 基于深度强化学习模型, 智能的动态选择具有高数据质量且能提高最终模型精度的客户端参与每一轮的联邦学习, 以抵消数据质量分布不平衡引入的偏差, 加速模型收敛并提高模型精度。在 MNIST 及 CIFAR-10 数据集上的实验表明, 所提出算法与其他算法相比, 在减少通信开销的同时进一步加快了收敛速度, 同时在模型最终准确性上也有较好的性能。

关键词: 联邦学习; 深度强化学习; 客户端动态采样; 贡献指数; 数据质量; 通信效率; 沙普利值; 模型精度
中图分类号: TP295 **文献标志码**: A **文章编号**: 1673−4785(2024)06−1552−10

中文引用格式: 赵泽华, 梁美玉, 薛哲, 等. 基于数据质量评估的高效强化联邦学习节点动态采样优化 [J]. 智能系统学报, 2024, 19(6): 1552–1561.
英文引用格式: ZHAO Zehua, LIANG Meiyu, XUE Zhe, et al. Client dynamic sampling optimization of efficient reinforcement federated learning based on data quality assessment[J]. CAAI transactions on intelligent systems, 2024, 19(6): 1552–1561.

Client dynamic sampling optimization of efficient reinforcement federated learning based on data quality assessment

ZHAO Zehua, LIANG Meiyu, XUE Zhe, LI Ang, ZHANG Min
(Beijing Key Laboratory of Intelligent Communication Software and Multimedia, Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract: Communication cost and efficiency are the key bottlenecks of federated learning due to the existence of system and statistical heterogeneities. Selecting only a subset of clients to perform model updates and aggregation can effectively reduce communication costs among numerous participants. However, biased selection and uneven distribution of data quality across clients pose additional challenges to client sampling methods. Therefore, this paper proposes a method for client dynamic sampling optimization in efficient reinforcement federated learning based on data quality assessment (RQCS) to address the aforementioned issues. This method evaluates data quality on clients using a contribution index based on the Shapley value and intelligently selects clients with high data quality for each round of federated learning. By leveraging reinforcement learning, the method aims to offset the bias introduced by uneven data quality distribution, accelerate model convergence, and improve model accuracy. Experiments on the MNIST and CIFAR-10 datasets show that the proposed algorithm not only reduces communication costs but also further accelerates convergence speed and achieves better performance in model accuracy compared to other algorithms.

Keywords: federated learning; deep reinforcement learning; client dynamic sampling; contribution index; data quality; communication efficiency; Shapley value; model accuracy

虽然大数据时代提供了海量数据, 但是由于隐私安全、法律法规和公司制度等问题, 导致大部分行业中的数据都是以孤岛形式存在[1]。联邦学习 (federated learning, FL) 正是面向这种数据孤

收稿日期: 2023−05−31. **网络出版日期**: 2024−09−12.
基金项目: 国家自然科学基金项目(62192784, U22B2038, 62172056, 62272058); 中国人工智能学会−华为 MindSpore 学术奖励基金项目 (CAAIXSJLJJ-2021-007B).
通信作者: 梁美玉. E-mail: meiyu1210@bupt.edu.cn.

岛现实场景而设计的机器学习范式[2]。具体地说, 联邦学习中央服务器和多个参与客户端之间通过传递模型参数保证数据可以在不出本地的情况下完成联邦学习模型的训练, 即数据不动模型动, 保证了客户端数据的隐私与安全。与传统的分布式机器学习技术相比, 联邦学习具有 2 个独有的特征[3]。首先, 客户端分布广泛, 且运算能力、网络带宽和通信稳定性等性能指标存在多样化, 这称为系统异构性[4]。其次, 训练数据以非独立同分布的方式存储在客户端之间, 对模型聚合产生了负面影响, 这称为统计异构性[4]。系统异构性和统计异构性造成了较大的通信开销和较低的通信效率。因此, 通信开销和通信效率成为了联邦学习的关键瓶颈之一, 研究如何降低联邦学习的通信开销变得十分重要[5]。

在优化通信开销的众多方法中, 本研究主要关注的是基于客户端采样的优化方法。因为在联邦学习中, 参与的客户端数量可能非常庞大, 但由于模型分发和重新上传的带宽受限, 在所有参与设备上并行执行模型更新和聚合是不切实际的, 所以一般只选取一部分客户端参与联邦学习的训练过程[5]。因此, 客户端采样方法对于降低联邦学习的通信开销, 提高联邦训练过程中的收敛速度和最终模型精度等至关重要。现有的客户端采样方法主要包括随机采样、基于客户端的数据数量选择客户端子集和基于梯度的相似性度量对客户端进行分层聚类。但上述方法仍然存在各种各样的问题, 比如: 随机选择的客户端的数据分布可能无法反映全局视图中的真实数据分布, 对特定客户端的选择过多可能会使全局模型“漂移”到其本地优化器, 导致对全局模型更新的偏差, 从而出现客户端“漂移”现象[6]; 使用聚类算法选择客户端虽然可以帮助平衡数据的分布, 但是在聚类过程中产生了大量的通信开销; 简单地将数据量作为评判客户端质量的指标, 但数据量大的客户端可能数据质量较低, 此时根据数据量选择客户端反而会降低模型质量。

总结以上方法中存在的问题: 1)随机选择中潜在的客户端“漂移”现象可能会损害模型性能或最终模型精度; 2)客户端采样算法(例如聚类采样)会增加额外的通信开销; 3)粗略地将样本量大小作为客户端质量的评估指标, 根据样本量大小选择客户端。针对以上问题本研究提出了基于数据质量评估的高效强化联邦学习节点动态采样优化方法(client dynamic sampling optimization of efficient reinforcement federated learning based on data quality assessment, RQCS): 采用沙普利值(shapley value)的贡献指数(contribution index, CI)[7]评估客户端上的数据质量, 并将贡献指数作为客户端采样的重要指标; 为了缓解客户端“漂移”现象, 将数据质量(贡献指数)和深度强化学习结合起来, 智能的动态选择拥有高数据质量且能提高模型性能和最终模型精度的客户端; 根据模型聚合过程中产生的中间结果计算贡献指数, 没有增加额外的通信负担。

本研究的主要贡献包括 4 个方面:

1) 不同于传统基于样本量的客户端采样方法, 首次提出了基于数据质量的客户端动态采样算法, 通过选择高数据质量的优质联邦节点参与模型聚合, 提高了强化联邦学习模型性能。

2) 首次将沙普利值的贡献指数应用到联邦学习的客户端采样问题中, 基于贡献指数评估客户端的数据质量, 并在大量联邦参与方中筛选出高数据质量的客户端。

3) 首次将贡献指数和深度强化学习相结合, 设计了奖励函数以综合考虑模型性能和模型精度, 提出了一种基于数据质量评估的高效强化联邦学习节点动态采样优化方法, 通过动态选取高数据质量的客户端有效降低了通信开销。

4) 在 2 个标准数据集 MNIST 和 CIFAR-10 上验证了所提方法的性能。实验结果表明, 本研究所提出的方法能够提高最终模型精度, 并使模型达到特定精度值所需的通信轮次更少, 效果优于其他的基线方法。

1　相关工作

近年来许多研究者在联邦学习中的客户端选择方面做了大量工作, 下面对现有的客户端选择方案进行介绍。

最为经典的是 Mcmahan 等[8] 提出的 FedAvg 算法, 该算法随机选择客户端参与联邦学习的训练, 但是其在非独立同分布数据下收敛速度较慢。随后 Li 等[9] 基于 FedAvg 提出了改进的 FedProx 算法, 在目标函数中添加了一个近端项来解决统计异质性问题, 在高度异构的环境中, FedProx 表现出比 FedAvg 更稳定和更准确的收敛行为。Nishio 等[10] 提出了 FedCS 算法, 解决了在资源受限情况下(如计算资源有限、无线信道条件较差)的客户端选择问题, 提高了整个训练过程的效率, 并减少了训练联邦学习模型所需的时间。Cho 等[11] 首次对有偏客户端选择策略进行了联邦优化的收敛

分析，研究发现将客户端选择偏向于具有较高局部损失的客户端可以实现更快的错误收敛，据此提出了 Power-of-Choice 客户端选择框架，在收敛速度和选择偏差之间取得了权衡。Fraboni 等[12]提出了基于客户端的样本数量和基于梯度的相似性度量 2 种聚类采样方法，虽然基于聚类选择客户端提高了数据样本的代表性，但是因为需要所有客户端向服务器发送梯度，在模型聚合期间会产生大量的通信开销。

此外，还有一些客户端选择方法不仅降低了通信开销，同时也考虑了数据的非独立同分布[13]特征。例如，Kollias 等[14]考虑了非独立同分布数据的分布特征，提出了基于草图和聚类的联邦学习客户端选择方法，该方法与 Fraboni 等[12]提出的聚类采样类似，但是 Kollias 等[14]使用了模型参数的草图，这在聚类步骤中产生了非常低的通信开销。Wang 等[15]基于强化学习提出了 Favor 方法，强化学习智能体通过在每轮通信中主动选择客户端设备的最佳子集来加速收敛，减少通信开销，并且该子集可以抵消非独立同分布数据引入的偏差。Huang 等[16]的研究表明，倾向性地选择稳定的客户端可能会增加有效参与而使得收敛速度加快，但选择偏差可能会使所获得的模型性能下降，因此重点研究选择的公平性，提出了一种客户端选择方案 E3CS 在客户端选择的稳定性和选择偏差之间取得了较好的权衡。为了提高采样过程中的动态性，Ji 等[17]提出了一种基于指数退火的联邦平均动态采样策略，使用指数衰减率来退火训练过程中的采样率，由动态变化的采样率动态控制所选客户端模型的比例。Rai 等[18]根据客户端上的数据量、本地类别不平衡以及是否非独立同分布等条件定义了无关性分数，在无关性分数的基础上提出了无关采样技术，缓解了搭便车问题对模型性能的影响。但是上述方法中没有考虑到客户端上的数据质量对模型性能的影响。本研究提出的基于数据质量评估的高效强化联邦学习节点动态采样优化方法，创新性地将数据质量作为客户端选择的重要指标，不仅缓解了客户端上的数据样本分布不平衡现象，而且选择高数据质量的训练样本能够显著提高模型的性能。为了解决选择偏差导致的客户端“漂移”问题，本研究将数据质量和深度强化学习机制有机结合，通过设计新的强化联邦奖励函数，在选择高数据质量客户端的同时也考虑该客户端对模型准确度的影响，智能动态选择能够提升模型性能和提高模型最终准确度的联邦参与设备。

2　算法设计

由图 1 可知，联邦学习通过迭代聚合来自多个客户端设备的模型梯度来训练共享的全局模型[19]，这些客户端设备可能拥有不同质量的数据集。为此，基于贡献指数及深度强化学习（deep reinforcement learning, DRL），针对联邦学习中存在的通信开销大，参与方数据质量分布不平衡等问题，提出了基于数据质量评估的高效强化联邦学习节点动态采样优化方法（RQCS）。首先将联邦学习客户端选择问题定义为基于深度强化学习的联邦节点动态采样问题，其次介绍了基于贡献指数的联邦节点数据质量评估方法，然后阐述了深度强化联邦学习的模型设计，接着介绍了数据质量引导的强化联邦学习节点动态采样优化模块，随后详细介绍了基于 RQCS 算法的强化联邦学习工作流程，最后介绍了基于双深度 Q 网络的强化联邦学习模型训练过程。

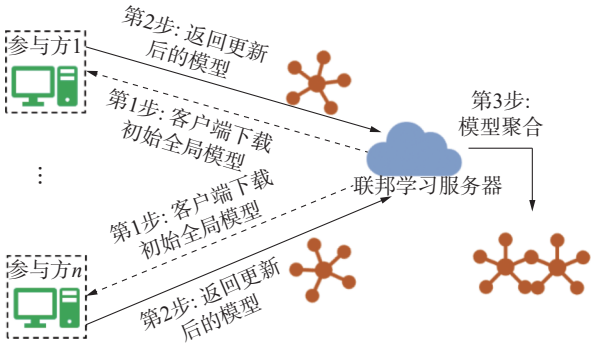


图 1　联邦学习架构

Fig. 1　Architecture of federated learning

2.1　问题定义

本研究将联邦学习客户端选择问题定义为基于深度强化学习的联邦节点动态采样问题。假设在n个数据提供者（客户端设备）上有一个联邦学习任务，目标精度为Γ，每个数据提供者的数据集表示为$D_i, i \in \mathrm{N} = \{1,2,\cdots,n\}$，$|D_i|$表示训练数据集$D_i$的大小，联邦学习的迭代轮数表示为$t \in \{0,1,\cdots,T\}$，其中，$T$代表联邦学习训练结束时迭代的总轮数。第$t$轮中客户端$i$上传的模型权重定义为$\{W_i^{(t)} | i \in \mathrm{N}\}$，用$W^{(t)}$表示第$t$轮的全局初始模型权重；同理$\{\Delta_i^{(t)} | i \in \mathrm{N}\}$表示客户端$i$在第$t$轮的梯度信息，$\Delta^{(t)}$表示第$t$轮的聚合梯度。为便于准确评估客户端$i$的数据质量，本研究用$M \subseteq \mathrm{N} = \{1,2,\cdots,n\}$表示$N$的子集，用$\mathrm{Per}(W_{\{i\}})$表示在客户端$i$的数据集上评估的模型性能。每轮联邦学习过程可以建模为马尔可夫决策过程[20]（Markov decision process, MDP），状态s由全局模型权重和每轮通信回合中各个客户端设备的模型权重表示。给定一个

状态, DRL 智能体采取一个动作a来选择K个客户端执行本地训练和更新全局模型, 然后观察奖励信号r, 这是全局模型在验证集上实现的测试精度v和客户端i的贡献指数ϕ_i的函数, 目标是训练 DRL 智能体基于贡献指数ϕ_i动态地选择高数据质量的客户端, 使强化联邦学习模型加快收敛速度并提高模型最终准确性。

2.2　基于贡献指数的联邦节点数据质量评估

为了计算每轮通信回合中各个客户端的贡献指数, 在每一轮训练中, 强化联邦学习服务器首先计算各个客户端上的梯度信息$\{\Delta_i^{(t)}\}_{i=1,2,\cdots,n}$, 通过各个客户端上样本量的加权平均来聚合这些梯度信息得到聚合梯度$\Delta_M^{(t)}$。然后根据各个客户端的聚合梯度近似地重建这些模型, 得到近似的全局模型$\tilde{W}_M^{(t+1)}$。对每一个客户端$i, i\in 1,2,\cdots,n$, 基于其训练数据评估其近似的全局模型的性能, 通过评估这些重建模型的性能来评估不同数据提供者的贡献指数。例如对 2 个客户端i和$j(j\in 1,2,\cdots,n)$上的训练样本D_i和D_j对模型的性能有相同的影响, 即$\mathrm{Per}(W_{M\cup\{i\}})=\mathrm{Per}(W_{M\cup\{j\}})$, 则其应该具有相同的贡献指数$\phi_i=\phi_j$。客户端$i$的贡献指数计算公式为

$$\phi_i = C\cdot\sum_{M\subseteq \mathrm{N}\setminus\{i\}}\frac{\mathrm{Per}(W_{M\cup\{i\}})-\mathrm{Per}(W_M)}{\binom{n-1}{|M|}} \tag{1}$$

式中C是一个常数。根据式 (1) 即可估计在当前训练轮次t中每个客户端的贡献指数$\phi_i^{(t+1)}$。

2.3　深度强化联邦学习模型设计

本研究中的深度强化联邦学习模型基于深度强化学习来优化联邦学习客户端动态采样算法, 提升联邦学习的质效均衡性。在介绍强化联邦学习的模型设计之前, 首先介绍一下强化学习的一些基础理论。

强化学习[21] 主要是智能体(agent)通过与环境的交互找到最佳的动作(action)以获得长期最大化回报(reward)的过程。交互过程可以通过一个 5 元组(S,A,R,V,γ)建模为 MDP[20], 其中, S表示状态集; A表示动作集; R是奖励函数, 其将每个状态$s\in S$和在状态s下采取的行动a映射到期望的即时奖励$r_t=R(s_t,a_t)$, $V(s,a)$是状态转换概率; $\gamma\in[0,1]$是反映当前奖励对未来奖励的重要性递减的折扣因子。MDP 的目标是找到一个策略$\pi^*(a|s)$(π^*表示为最优策略)来决定在状态s下选择的动作a, 从而最大化智能体的期望累积奖励$R=\sum_{t=1}^{T}\gamma^{t-1}r_t$。从状态$s_t$到状态$s_{t+1}$的一系列转换过程产生的预期累积奖励可以通过贝尔曼方程[22]定义为动作−价值函数[23]:

$$V_\pi(s_t,a)=E_{s_{t+1},a}[r_t+\gamma V_\pi(s_{t+1},a)|s_t,a_t] \tag{2}$$

式中: π是从状态到采取动作的一个映射, π^*和最优动作−价值函数$V^*(s_t,a)$对应; $E_{s_{t+1},a}$表示从状态s_t到状态s_{t+1}通过采取行动a的一系列转换过程而产生的预期累积奖励; r_t表示即时奖励。最优价值函数公式为

$$V^*(s_t,a)=E_{s_{t+1}}\left[r_t+\gamma\max_a V^*(s_{t+1},a)|s_t,a\right] \tag{3}$$

那么在每次状态转换中, $r_t+\gamma\max_a V(s_{t+1},a;\tau_t)$就成为了$V(s_t,a;\tau_t)$学习的目标。通常用深度神经网络来表示最优动作−价值函数的逼近器, 强化学习问题就转化为最小化目标和逼近器之间的均方误差损失[24], 定义为

$$\vartheta_t(\tau_t)=\left(r_t+\gamma\max_a V(s_{t+1},a;\tau_t)-V(s_t,a;\tau_t)\right)^2 \tag{4}$$

式中$\vartheta_t(\tau_t)$表示均方误差损失。

根据以上理论基础, 本研究确定了基于强化联邦学习环境的深度 Q 网络(deep q-learning network, DQN)模型中的状态空间和动作空间。假设在n个数据提供者(客户端设备)上有一个联邦学习任务, 目标精度为Γ。为了提高训练效率, 本研究将强化联邦学习服务器视为基于 DQN 的智能体, 负责与环境交互并在每一轮中动态地选择K个数据提供者(客户端)参与训练。系统状态实时更新并在客户端之间共享, 智能体仅收集服务器和客户端的模型权重信息作为状态, 不需要搜集或检查任何数据样本, 从而维持联邦学习的隐私保护级别。之后, 智能体通过构建系统状态和选择最优动作来做出自适应客户端选择的决策。此问题中的目标是训练深度强化学习智能体尽快收敛到目标精度。

状态空间　由于在联邦学习的训练过程中, 模型权重将在每一轮通信结束时更新, 因此, 本研究根据第t轮时服务器上的模型权重$W^{(t)}$和客户端上的模型权重$\{W_1^{(t)},W_2^{(t)},\cdots,W_n^{(t)}\}$定义第$t$轮中的系统状态为$s_t=(W^{(t)},W_1^{(t)},\cdots,W_n^{(t)})$。联邦学习服务器上的智能体维护着模型权重列表$\{W_i|i\in\mathrm{N}\}$, 其中$\mathrm{N}=\{1,2,\cdots,n\}$, 并且仅当设备$i$在第$t$轮被选中参与训练得到新的梯度信息$\Delta_i^{(t)}$时$W_i$才会更新。

动作空间　在本研究的问题中, DQN 智能体负责在每个通信回合t开始时采取行动从全部的N个数据提供者(客户端)中动态地选择K个数据提供者(客户端)参与训练。实际上基于 DQN 的智能体在每轮训练过程中仅从N个数据提供者(客户端)中选择一个客户端, 而在测试和应用

中，智能体将对一批排名前K的客户端进行采样，使之参与到联邦学习的训练过程。因此，基于DQN的智能体需要学习到最优的动作价值函数$V^*(s_t,a)$，即估计从状态s_t开始的能使预期回报最大化的动作。因此动作空间为$\{1,2,\cdots,n\}$，其中$a=i$意味着客户端i被选中参与联邦学习的训练过程。一旦DQN在测试期间被训练为近似的最优价值函数$V^*(s,a)$，那么在第t轮中，DQN代理将为所有的N个动作计算对应的最优价值函数$\{V^*(s_t,a)|a\in \mathrm{N}\}$。每个动作−价值函数表示智能体在状态$s_t$下选择特定动作$a$可以获得的最大预期回报。最后通过选择$K$个设备，每个设备对应一个不同的动作$a$，从而产生$V^*(s_t,a)$的前$K$个值。

2.4　数据质量引导的强化联邦学习节点动态采样优化

提出数据质量引导的高效强化联邦学习节点动态采样优化策略，主要体现在强化联邦学习环境中的奖励函数设计部分，下面将主要介绍奖励函数的设计。

本研究将每个通信回合t结束时的奖励设置为$r_t=\left(1+\phi_i^{(t)}\right)^{(v_t-\Gamma)}-1$，$t=1,2,\cdots,T$，其中，$v_t$是全局模型在第$t$轮后在保留的验证集上实现的测试精度，$\Gamma$是目标精度，$\phi_i^{(t)}$是客户端$i$在第$t$轮中的贡献指数，即数据提供者$i$上的数据质量，$1+\phi_i$是贡献指数的正向激励，同时也确保了第$t$轮的奖励$r_t$随测试精度$v_t$呈指数增长。因为$0\leqslant v_t\leqslant\Gamma\leqslant 1$，$0\leqslant\phi_i^{(t)}\leqslant 1$，所以有$r_t\in(-0.5,1]$。当测试精度达到目标精度，即$v_t=\Gamma$时联邦学习训练停止，此时$r_t$达到其最大值0。

通过训练基于DQN的智能体来最大化期望的累积折扣奖励R，其表达式为

$$R=\sum_{t=1}^{T}\gamma^{t-1}r_t=\sum_{t=1}^{T}\gamma^{t-1}\left(\left(1+\phi_i^{(t)}\right)^{(v_t-\Gamma)}-1\right)\tag{5}$$

式中，$\gamma\in(0,1]$是对未来回报的折扣因子，避免总是使得当前回报最大化而忽略长期回报。

式(5)中的第1项激励智能体选取贡献指数较高（高数据质量）的客户端并实现更高的测试准确性，$1+\phi_i^{(t)}$控制奖励r_t随v_t的增长速度；式(5)中的第2项的−1是为了鼓励以更少的轮数完成联邦学习训练，因为需要的轮数越多，智能体获得的累积奖励就越少。

2.5　基于RQCS算法的强化联邦学习工作流程

图2为本研究提出的联邦节点动态采样优化方法RQCS如何基于DRL智能体在每一轮中动态选择客户端设备进行联邦学习模型训练。算法1和算法2总结了基于RQCS的强化联邦学习算法的详细流程，其中详细步骤如下。

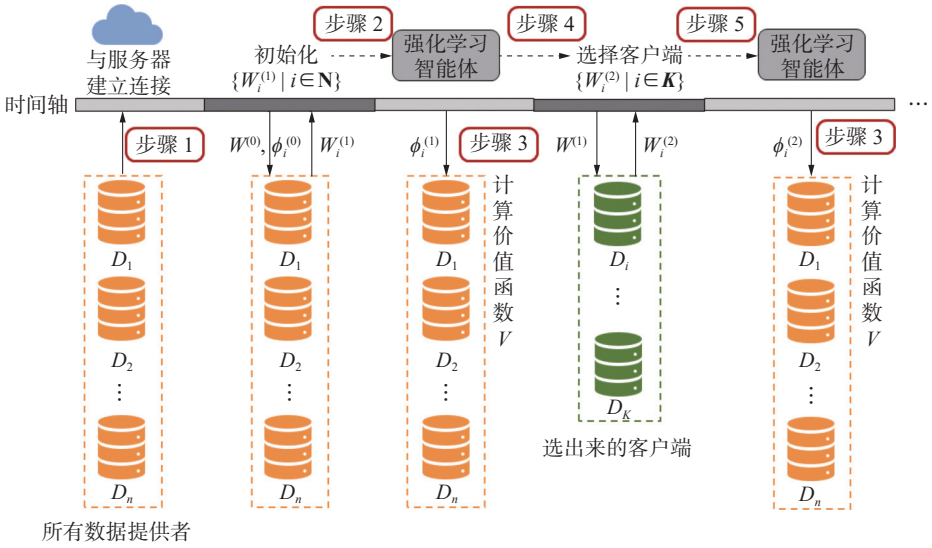


图 2　使用 RQCS 算法的联邦学习流程

Fig. 2　Federated learning workflow with RQCS

1) 所有N个客户端与联邦学习服务器建立连接，确保设备可用。

2) 每个客户端都从服务器下载初始的全局模型权重$W^{(0)}$，使用M来表示集合$\{1,2,\cdots,n\}$，并初始化基于不同非空子集$M\subseteq\mathrm{N}$的重建模型$\left\{\tilde{W}_M^{(0)}|M\subseteq\mathrm{N}\right\}$，从而得到初始化的贡献指数$\phi_i^{(0)}$。

3) 在第t轮中，其中，$t=1,2,\cdots,T$，首先计算客户端的梯度$\{\Delta_i^{(t)}\}_{i=1,2,\cdots n}$进行梯度聚合（算法1第4行），算法1第5~7行表示根据来自客户端的梯度近似地重建每轮的初始全局模型，而不是在N的所有非空子集上重新训练这些模型。9~11行表示计算不同客户端（数据提供者）在当前通信

回合中的贡献指数。然后基于 DQN 的智能体计算所有设备的$V(s_t,a;\tau)$。

4) 第 13~14 行表示基于 DQN 的智能体根据计算的$V(s_t,a;\tau)$选择前K个客户端, 被选中的K个客户端会在本地执行P个 epoch 的本地随机梯度下降并得到$\{W_i^{(t+1)}|i\in K\}$。

5) $\{W_i^{(t+1)}|i\in K\}$被上传到服务器进行模型聚合, 计算全局模型$W^{(t+1)}$。进入第$t+1$轮并重复步骤 3)~5)。

重复步骤 3)~5) 直到达到目标精度或达到一定的通信回合。

算法 1　基于 RQCS 的强化联邦学习算法

输入　每轮选择的客户端数量K。

输出　全局模型$W^{(t+1)}$。

1) $\mathrm{N}\leftarrow\{1,2,\cdots n\}$;

2) 初始化$W^{(0)},\{\tilde{W}_M^{(0)}|M\subseteq\mathrm{N}\},\phi_i^{(0)}$;

3) for 每一轮$t\leftarrow 1,2,\cdots,T$执行;

4) $\Delta_i^{(t)}\leftarrow W_i^{(t)}-W^{(t)}$对客户端$i\in\mathrm{N}$;

5) for 每个子集$M\subseteq\mathrm{N}$并行执行;

6) $\Delta_M^{(t)}\leftarrow\sum_{i\in\mathrm{N}}\frac{|D_i|}{\sum_{i\in\mathrm{N}}|D_i|}\cdot\Delta_i^{(t)}$;

7) $\tilde{W}_M^{(t+1)}\leftarrow\tilde{W}_M^{(t)}+\Delta_M^{(t)}$;

8) end for;

9) for $i\leftarrow 1,2,\cdots,n$ 并行执行;

10) $\phi_i^{(t+1)}=C\cdot\sum_{M\subseteq\mathrm{N}\setminus\{i\}}\left(\left(\mathrm{Per}(\tilde{W}_{M\cup\{i\}}^{(t+1)})-\mathrm{Per}(\tilde{W}_M^{(t+1)})\right)/\binom{n-1}{|M|}\right)$;

11) end for;

12) $V(s_t,a;\tau)\leftarrow$计算价值函数;

13) agent 根据价值函数选择 top-K 设备;

14) $W_i^{(t+1)}\leftarrow\mathrm{ClientUpdate}(i,W^{(t)}),i\in K$;

15) $W^{(t+1)}\leftarrow\sum_{i\in K}\frac{|D_i|}{\sum_i|D_i|}\cdot W_i^{(t+1)}+\sum_{i\in\mathrm{N}\wedge i\notin K}\frac{|D_i|}{\sum_i|D_i|}\cdot W^{(t)}$;

16) end for;

17) 返回$W^{(t+1)}$。

算法 2　$\mathrm{ClientUpdate}(i,W)$

输入　本地批量大小 B, 本地迭代次数 P, 学习率 η;

输出　局部模型W。

1) $\beta\leftarrow$将数据集D_i划分成B个批量大小;

2) for 每次本地迭代$p\leftarrow 1,2,\cdots,P$执行;

3) for 批量$b\in\beta$并行执行;

4) $W\leftarrow W-\eta\nabla L(W;b)$;

5) end for;

6) end for;

7) 返回W给服务器。

2.6　基于双深度 Q 网络的强化联邦学习模型训练

本研究基于双深度 Q 网络(double deep q-learning network, DDQN)学习强化联邦学习模型的最优价值函数$V^*(s_t,a)$。Q 学习算法根据选择的客户端设备为状态s_t处的每个潜在动作a提供价值估计。但原始的 Q 学习算法可能不稳定, 因为其通过不断逼近$V(s_t,a;\tau_t)$学习到最佳的动作−价值函数$V^*(s_t,a)$。DDQN 增加了另一个值函数$V(s,a;\tau_t')$来稳定动作−价值函数的估计。DDQN 的设计动机是网络每x次更新和冻结一次。DDQN 增加了动作−价值评估的稳定性, 避免了“抖动”问题。

为了训练 DRL 智能体, 强化联邦学习服务器在第1轮时执行随机客户端设备选择以初始化状态。状态被反馈到 DDQN 之一的 Q 网络$V(s_t,a;\tau_t)$。DDQN 生成一个动作a来为联邦学习服务器选择一个客户端设备。经过几轮联邦学习训练后, DRL 智能体已经采样了一些动作−状态对, 智能体通过以下公式解决式 (4) 中的问题:

$$\vartheta_t(\tau_t)=\left(Y_t^{\mathrm{DoubleQ}}-V(s_t,a;\tau_t)\right)^2 \tag{6}$$

其中第t轮的目标Y_t^{DoubleQ}定义为

$$Y_t^{\mathrm{DoubleQ}}=r_t+\gamma\max_a V(s_{t+1},a;\tau_t) \tag{7}$$

$$Y_t^{\mathrm{DoubleQ}}=r_t+\gamma V\left(s_t,\arg\max_a V(s_t,a;\tau_t);\tau_t'\right) \tag{8}$$

式 (8) 使用 2 个动作−价值函数来更新Y_t^{DoubleQ}, 其中, τ_t是每个时间步骤中更新的在线参数, τ_t'是冻结参数, 用来增加动作−价值估计的稳定性。通过梯度下降最小化$\vartheta_t(\tau_t)$来更新动作−价值函数$V(s_{t+1},a;\tau_t)$。

$$\tau_{t+1}=\tau_t+\alpha(Y_t^{\mathrm{DoubleQ}}-V(s_t,a;\tau_t))\nabla_{\tau_t}V(s_t,a;\tau_t) \tag{9}$$

式中α是步长。

3　实验结果与分析

为验证本研究所提算法 RQCS 的性能, 首先描述了实验数据集和评价指标、对比方法和实验设置, 然后将提出的 RQCS 和已有的其他联邦学习客户端采样算法进行了实验对比。

3.1　数据集和评价指标

本研究在 2 个联邦学习标准数据集上分别验证了提出的算法性能, 分别是手写数字识别数据集 MNIST 和图像分类数据集 CIFAR-10。对于 MNIST 数据集上的手写数字识别任务, 本研究使用的是由 2 个5×5卷积层(每个卷积层的通道数为 10)、1 个2×2最大池化层、2 个分别具有 1 280 和 256 个单元的全连接层和 1 个 softmax 输出层

组成的卷积神经网络。对于具有挑战性的 CIFAR-10 数据集的图像分类任务，本研究使用另一个由 2 个5×5卷积层（每个卷积层的通道数为64）、1 个2×2最大池化层、2 个分别具有 384 和 192 个单元的全连接层和 1 个 softmax 输出层组成的卷积神经网络。

本研究的评价指标为通信轮数和测试精度 2 个变量，通过控制变量法对算法性能进行评估。在控制测试精度相同的情况下，比较联邦学习框架中采用不同的客户端选择算法时模型聚合所需的通信轮数；在控制通信轮数相同的情况下，比较联邦学习框架中采用不同的客户端选择算法时模型达到的测试精度。

3.2　对比方法

本研究将 RQCS 与 4 种最先进的方法进行比较，Random[8]、FedCS[10]、pow-d[11] 和 Favor[15]。Random 随机选择客户端参与联邦学习的训练；FedCS 允许服务器聚合尽可能多的客户端更新的模型以在更短的时间内提升模型的性能；pow-d 偏向选择具有较高局部损失的客户端以实现更快的收敛；Favor 基于强化学习选择客户端设备，以抵消非独立同分布 (non-iid) 数据引入的偏差并加速收敛。

3.3　实验设置

本研究通过独立同分布 (independent identically distribution, IID) 设置和非独立同分布 (Non-IID) 设置模拟数据分布。对于 iid 设置，每个客户端独立抽样数据集中的数据片段。对于 Non-IID 设置，本研究为每个客户端随机选择 1 个主标签，然后从数据集中采样 80% 与其主标签一致的数据片段，再从剩下的标签中采样 20%的数据片段。对于 IID 设置和 Non-IID 设置，每个客户端随机保留 10% 的数据用于测试。

本研究实验环境采用深度学习框架 PyTorch，在高性能工作站 (12 vCPU Intel(R) Xeon(R) Platinum 8255C CPU @ 2.50 GHz, RTX 2080 Ti(11 GB) * 1) 上模拟了联邦学习的训练过程，模拟对象包括 1 台服务器和 100 个客户端。

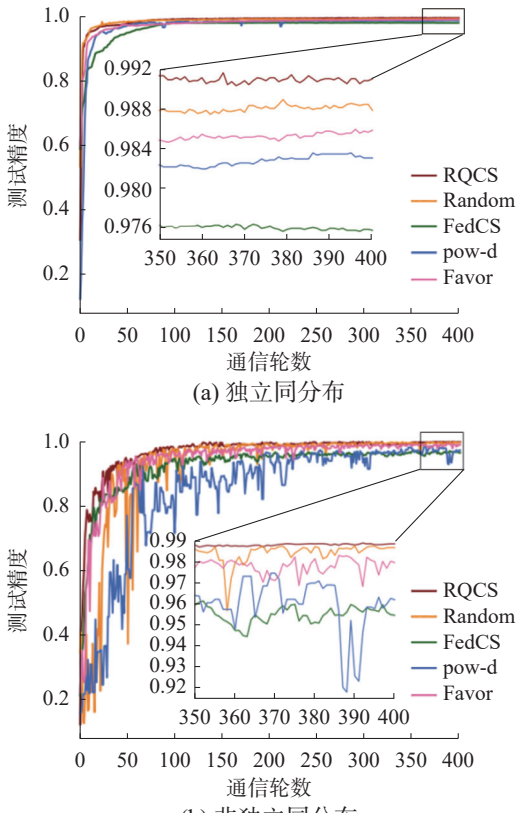
在实验中发现，在联邦学习的开始，由于模型是随机初始化的，因此全局模型可能表现不佳，一些数据提供者可能有负置信区间。因此，在前几轮训练中，一旦某个客户端的贡献指数为负，本研究就采用加权平均的思想，为每个数据提供者分配相同的贡献指数。

为了验证本研究所提算法 RQCS 的有效性，将其与随机客户端选择算法（Random）[8]、FedCS[10]、pow-d[11] 和 Favor[15] 做了对比，本研究控制 5 种算法的实验条件基本相同，包括参与者之间的数据划分、全局模型初始化数值和一些参数设置。实验中，设置每轮采样 K=10 个客户端，本地批量大小 batch_size=64，本地迭代次数 epochs=4，学习率 $\eta = 0.01$。针对不同的客户端选择方案在 MNIST 数据集和 CIFAR-10 数据集上分别测试了在独立同分布和非独立同分布设置下进行 400 轮和 2 500 轮通信，以评估收敛速度和实际训练中的最终模型性能。

3.4　实验结果与分析

3.4.1　不同方法的收敛速度分析

在 MNIST 数据集上，数据是独立同分布的情况下，5 种算法经过 400 轮聚合后的实验结果如图 3(a) 所示，通过比较图 3(a) 中的收敛曲线可以发现，RQCS 的收敛速度仅次于 Random，Favor 的收敛速度次之，pow-d、FedCS 的收敛速度最慢。表 1 给出了在模型准确率分别达到 90%、95% 和 97% 时，5 种算法所需的通信轮次，通过观察表 1 可以得出同样的结论。以 95% 的模型准确率作为目标精度时，RQCS 的通信轮次最多能减少 75%。这是因为 MNIST 数据集上的手写数字识别任务相对简单，并且数据是独立同分布的，在此环境下，随机客户端选择方法不会出现客户端“漂移”现象，所以 RQCS 的收敛速度稍慢于 Random。



(a) 独立同分布

(b) 非独立同分布

图 3　MNIST 数据集上通信轮数和测试精度的关系

Fig. 3　Test accuracy vs. communication rounds for MNIST

表 1　MNIST 数据集，独立同分布
Table 1　Dataset of MNIST, IID　%

算法	准确率@90	准确率@95	准确率@97	最终准确率
Random	4	11	24	98.89
FedCS	29	57	87	97.67
pow-d	14	34	59	98.35
Favor	8	25	76	98.59
RQCS(本文)	5	14	46	99.14

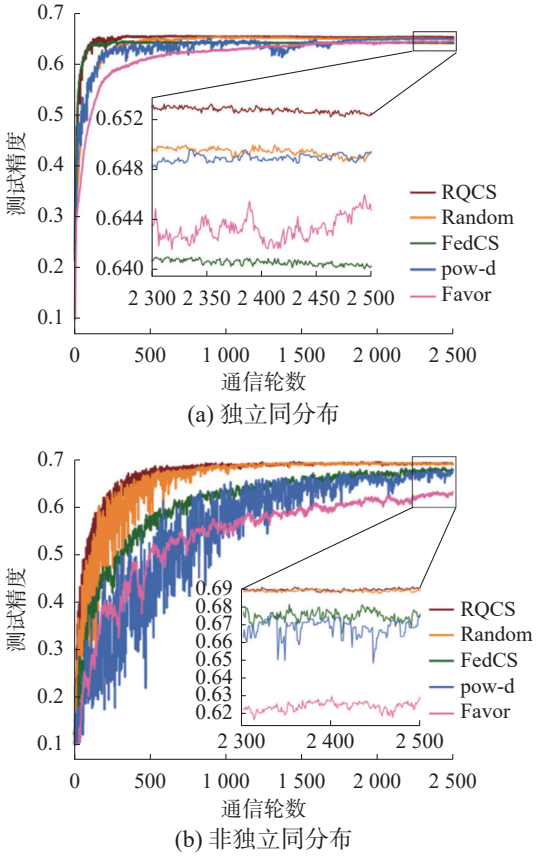
在 MNIST 数据集上, 数据是非独立同分布的情况下, 5 种算法经过 400 轮聚合后的实验结果如图 3(b) 所示, 通过比较图 3(b) 中的收敛曲线可以发现, 5 种算法的收敛曲线均出现了波动, 但 RQCS 的波动性最小, 说明即使是在非独立同分布的情况下, RQCS 相比于其他 4 种基线方法仍能稳定地收敛。通过观察表 2 发现, 在模型准确率分别达到 85%、90% 和 95% 时, 5 种算法中 RQCS 所需的通信轮次均是最少的。综合图 3(a) 和表 2 的结果可以证明, RQCS 基于数据质量评估进行客户端采样可以解决客户端上的数据质量分布不平衡问题, 并以此抵消非独立同分布数据引入的偏差, 从而使强化联邦学习模型快速稳定地收敛。

表 2　MNIST 数据集，非独立同分布
Table 2　Dataset of MNIST, Non-IID　%

算法	准确率@85	准确率@90	准确率@95	最终准确率
Random	34	40	74	98.76
FedCS	52	56	172	95.97
pow-d	59	67	196	97.46
Favor	40	45	82	98.28
RQCS(本文)	24	31	66	98.89

在 CIFAR-10 数据集上, 数据是独立同分布的情况下, 5 种算法经过 2 500 轮聚合后的实验结果如图 4(a) 所示, 通过比较图 4(a) 中的收敛曲线可以发现, 相比于其他 4 种基线方法, RQCS 的收敛速度最快。观察表 3 得知, 在模型准确率达到 60% 时, RQCS 的通信轮次最多能减少 87%。

在 CIFAR-10 数据集上, 数据是非独立同分布的情况下, 5 种算法经过 2 500 轮聚合后的实验结果如图 4(b) 所示, 通过比较图 4(b) 中的收敛曲线可以发现, 5 种算法的收敛曲线均出现了波动, 但 RQCS 的波动性最小, 说明即使是对于比较复杂的任务, 且数据是非独立同分布的情况下, RQCS 相比于其他 4 种基线方法仍能稳定地收敛。通过观察表 4 发现, 在模型准确率分别达到 50%、55% 和 60% 时, 5 种算法中 RQCS 所需的通信轮次均是最少的。综合图 4(b) 和表 4 的结果, 进一步证明了 RQCS 可以解决客户端上的数据质量分布不平衡问题, 并以此抵消非独立同分布数据引入的偏差, 从而使强化联邦学习模型快速稳定地收敛。



(a) 独立同分布

(b) 非独立同分布

图 4　CIFAR-10 数据集上通信轮数和测试精度的关系
Fig. 4　Test accuracy vs. communication rounds for CIFAR-10

表 3　CIFAR-10 数据集（独立同分布）
Table 3　Dataset of CIFAR-10, IID　%

算法	准确率@50	准确率@55	准确率@60	最终准确率
Random	57	89	166	65.06
FedCS	18	25	48	64.15
pow-d	41	86	148	64.92
Favor	107	160	368	64.59
RQCS(本文)	17	27	46	65.38

表 4　CIFAR-10 数据集（非独立同分布）
Table 4　Dataset of MNIST, Non-IID　%

算法	准确率@50	准确率@55	准确率@60	最终准确率
Random	119	150	241	68.92
FedCS	306	401	659	67.78
pow-d		559		67.43
Favor	501	781	1 587	62.92
RQCS (本文)	85	114	190	69.12

3.4.2　模型最终性能分析

通过观察图 3 和图 4 中放大的实验结果图可以发现，对于 MNIST 数据集上的手写数字识别任务和 CIFAR-10 数据集上具有挑战性的图像分类任务，无论数据是否是非独立同分布，RQCS 的模型最终准确率均高于其他 4 种基线方法，进一步观察表 1~4 可知，在 MNIST 数据集上，数据是独立同分布的情况下，相比其他 4 种基线方法，RQCS 的模型最终准确率最高提升了 1.47%；在 MNIST 数据集上，数据是非独立同分布的情况下，相比其他 4 种基线方法，RQCS 的模型最终准确率最高提升了 2.92%；在 CIFAR-10 数据集上，数据是独立同分布的情况下，相比其他 4 种基线方法，RQCS 的模型最终准确率最高提升了 1.23%；在 CIFAR-10 数据集上，数据是非独立同分布的情况下，相比其他 4 种基线方法，RQCS 的模型最终准确率最高提升了 6.2%。这证明了 RQCS 能够提高模型最终性能，因为 RQCS 选择的是高数据质量的训练样本。最重要的是，这证明了本研究提出的 RQCS 算法中，深度强化学习奖励函数设计的有效性。

4　结束语

针对强化联邦学习节点采样中存在通信开销大、参与方数据质量分布不平衡等问题，本研究提出了一种基于数据质量评估的高效强化联邦学习节点动态采样优化方法（RQCS）。在每一轮联邦学习中，RQCS 首先通过聚合上一轮次中各个客户端的梯度信息重建初始全局模型，根据模型性能评估各个客户端的贡献指数，即评估各个数据提供者的数据质量，然后基于深度强化学习智能体将贡献指数和模型精度作为奖励项，将通信轮次作为惩罚项，通过训练双 DQN 网络学习到近似最优的动作−价值函数，最后，强化学习智能体根据近似最优的动作−价值函数选择前K个设备参与到联邦学习的训练过程中。通过与其他基线方法的实验对比，验证了本研究提出的方法 RQCS 的有效性。

参考文献：

[1] 林伟伟, 石方, 曾岚, 等. 联邦学习开源框架综述[J]. 计算机研究与发展, 2023, 60(7): 1551−1580.
LIN Weiwei, SHI Fang, ZENG Lan, et al. Survey of federated learning open-source frameworks[J]. Journal of computer research and development, 2023, 60(7): 1551−1580.

[2] 田家会, 吕锡香, 邹仁朋, 等. 一种联邦学习中的公平资源分配方案[J]. 计算机研究与发展, 2022, 59(6): 1240−1254.
TIAN Jiahui, LYU Xixiang, ZOU Renpeng, et al. A fair resource allocation scheme in federated learning[J]. Journal of computer research and development, 2022, 59(6): 1240−1254.

[3] LUO Bing, XIAO Wenli, WANG Shiqiang, et al. Tackling system and statistical heterogeneity for federated learning with adaptive client sampling[C]//IEEE INFOCOM 2022-IEEE Conference on Computer Communications. London: IEEE, 2022: 1739−1748.

[4] 李少波, 杨磊, 李传江, 等. 联邦学习概述: 技术、应用及未来[J]. 计算机集成制造系统, 2022, 28(7): 2119−2138.
LI Shaobo , YANG Lei , LI Chuangjiang, et al. Overview of federated learning: technology, applications and future [J]. Computer integrated manufacturing systems, 2022, 28(7): 2119−2138.

[5] 邱鑫源, 叶泽聪, 崔翛龙, 等. 联邦学习通信开销研究综述[J]. 计算机应用, 2022, 42(2): 333−342.
QIU Xinyuan, YE Zecong, CUI Xiaolong, et al. Survey of communication overhead of federated learning[J]. Journal of computer applications, 2022, 42(2): 333−342.

[6] KARIMIREDDY S P, KALE S, MOHRI M, et al. SCAFFOLD: stochastic controlled averaging for federated learning[EB/OL]. (2019−10−14)[2021−01−01]. http://arxiv.org/abs/1910.06378.

[7] SONG Tianshu, TONG Yongxin, WEI Shuyue. Profit allocation for federated learning[C]//2019 IEEE International Conference on Big Data. Los Angeles: IEEE, 2019: 2577−2586.

[8] MCMAHAN H B, MOORE E, RAMAGE D, et al. Communication-efficient learning of deep networks from decentralized data[EB/OL]. (2016−02−17)[2021−01−01]. http://arxiv.org/abs/1602.05629.

[9] LI Tian, SAHU A K, ZAHEER M, et al. Federated optimization in heterogeneous networks[EB/OL]. (2018−12−14)[2021−01−01]. http://arxiv.org/abs/1812.06127.

[10] NISHIO T, YONETANI R. Client selection for federated learning with heterogeneous resources in mobile edge [C]//ICC 2019—2019 IEEE International Conference on Communications. Shanghai: IEEE, 2019: 1−7.

[11] CHO Y J, WANG Jianyu, JOSHI G. Client selection in federated learning: convergence analysis and power-of-choice selection strategies[EB/OL]. (2020−12−03)[2021−01−01]. http://arxiv.org/abs/2010.01243.

[12] FRABONI Y, VIDAL R, KAMENI L, et al. Clustered sampling: low-variance and improved representativity for clients selection in federated learning[EB/OL]. (2021−

05−12)[2021−12−01]. http://arxiv.org/abs/2105.05883.

[13] ZHAO Yue, LI Meng, LAI Liangzhen, et al. Federated learning with non-IID data[EB/OL]. (2018−06−02)[2021−01−01]. http://arxiv.org/abs/1806.00582.

[14] KOLLIAS G, SALONIDIS T, WANG Shiqiang. Sketch to Skip and Select: communication efficient federated learning using locality sensitive hashing[M]//Lecture Notes in Computer Science. Cham: Springer International Publishing, 2023: 72−83.

[15] WANG Hao, KAPLAN Z, NIU Di, et al. Optimizing federated learning on non-IID data with reinforcement learning[C]//IEEE INFOCOM 2020-IEEE Conference on Computer Communications. Toronto: IEEE, 2020: 1698−1707.

[16] HUANG Tiansheng, LIN Weiwei, SHEN Li, et al. Stochastic client selection for federated learning with volatile clients[J]. IEEE internet of things journal, 2022, 9(20): 20055−20070.

[17] JI Shaoxiong, JIANG Wenqi, WALID A, et al. Dynamic sampling and selective masking for communication-efficient federated learning[EB/OL]. (2020−03−21)[2021−01−01]. http://arxiv.org/abs/2003.09603.

[18] RAI S, KUMARI A, PRASAD D K. Client selection in federated learning under imperfections in environment[J]. AI, 2022, 3(1): 124−145.

[19] LI Qinbin, HE Bingsheng, SONG D. Model-contrastive federated learning[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021: 10708−10717.

[20] WACHI A, SUI Yanan. Safe reinforcement learning in constrained Markov decision processes[EB/OL]. (2008−08−15)[2021−01−01]. http://arxiv.org/abs/2008.06626.

[21] HENDERSON P, ISLAM R, BACHMAN P, et al. Deep reinforcement learning that matters[C]//Proceedings of the AAAI Conference on Artificial Intelligence. AAAI , 2018.

[22] FENG Y, LI L, LIU Q. A kernel loss for solving the bellman equation[J]. Advances in neural information processing systems, 2019, 32.

[23] ZHANG Jie, GUO Song, QU Zhihao, et al. Adaptive federated learning on non-IID data with resource constraint [J]. IEEE transactions on computers, 2022, 71(7): 1655−1667.

[24] NICOLSON A, PALIWAL K K. Deep learning for minimum mean-square error approaches to speech enhancement[J]. Speech communication, 2019, 111: 44−55.

作者简介:

赵泽华, 硕士, 主要研究方向为联邦学习中的高效通信。E-mail: zehuazhao@bupt.edu.cn。

梁美玉, 教授, 博士生导师, 主要研究方向为人工智能、数据挖掘、多媒体信息处理、计算机视觉。主持和参与国家自然科学基金重大项目、国家重点研发计划项目、973 计划课题、国家自然科学基金重点项目/重大国际合作项目/面上项目/青年科学基金等科研项目。发表学术论文 100 余篇, 出版学术专著 3 部, 申请和授权发明专利 40 余项。E-mail: meiyu1210@bupt.edu.cn。

薛哲, 副教授。主要研究方向为人工智能、机器学习、数据挖掘、多媒体信息处理。E-mail: xuezhe@bupt.edu.cn。