

数字孪生辅助联邦学习中的边缘选择和资源分配联合优化

唐伦 文明艳* 单贞贞 陈前斌
(重庆邮电大学通信与信息工程学院 重庆 400065)
(移动通信技术重庆市重点实验室 重庆 400065)

摘要: 在基于联邦学习的智能驾驶中, 智能网联汽车(ICV)的资源限制和可能出现的设备故障会导致联邦学习训练精度下降、时延和能耗增加等问题。为此该文提出数字孪生辅助联邦学习中的边缘选择和资源分配优化方案。该方案首先提出数字孪生辅助联邦学习机制, 使得ICV能够选择在本地或利用其数字孪生体参与联邦学习。其次, 通过构建数字孪生辅助联邦学习的计算和通信模型, 建立以最小化累积训练时延和能耗为目标的边缘选择和资源分配联合优化问题, 并将其转化为部分可观测的马尔可夫决策过程。最后, 提出基于多智能体参数化Q网络(MPDQN)的边缘选择和资源分配算法, 用于学习近似最优的边缘选择和资源分配策略, 以实现联邦学习累积时延和能耗最小化。仿真结果表明, 所提算法在保证模型精度的同时, 有效降低联邦学习累积训练时延和能耗。

关键词: 智能驾驶; 联邦学习; 数字孪生; 深度强化学习

中图分类号: TN929.5

文献标识码: A

文章编号: 1009-5896(2024)04-1343-10

DOI: 10.11999/JEIT230421

Joint Optimization of Edge Selection and Resource Allocation in Digital Twin-assisted Federated Learning

TANG Lun WEN Mingyan SHAN Zhenzhen CHEN Qianbin

(School of Communications and Information Engineering, Chongqing University of Posts and
Telecommunications, Chongqing 400065, China)

(Chongqing Key Laboratory of mobile Communications Technology, Chongqing, 400065, China)

Abstract: In intelligent driving based on federated learning, the resource constraints of Intelligent Connected Vehicle (ICV) and possible device failures will lead to the decrease of the precision of federated learning training and the increase of delay and energy consumption. Therefore, an optimization scheme of edge selection and resource allocation in digital twin-assisted federated learning is proposed. Firstly, a digital twin-assisted federated learning mechanism is proposed, allowing ICV to choose to participate in federated learning locally or through its digital twin. Secondly, by constructing a computational and communication model for digital twin-assisted federated learning, an edge selection and computing resource allocation joint optimization problem is established with the objective of minimizing cumulative training delay and energy consumption, and is transformed into a partially observable Markov decision process. Finally, an edge selection and resource allocation algorithm based on Multi-agent Parametrized Deep Q-Networks (MPDQN) is proposed to learn approximately optimal edge selection and resource allocation strategies to minimize federated learning cumulative delay and energy consumption. Simulation results show that the proposed algorithm can effectively reduce cumulative training delay and energy consumption of federated learning training while ensuring model accuracy.

Key words: Intelligent driving; Federated learning; Digital twin; Deep reinforcement learning

收稿日期: 2023-05-15; 改回日期: 2023-09-14; 网络出版: 2023-09-15

*通信作者: 文明艳 wenming155968@163.com

基金项目: 国家自然科学基金(62071078), 重庆市教委科学技术研究项目(KJZD-M201800601), 四川省科技计划(2021YFQ0053)

Foundation Items: The National Natural Science Foundation of China (62071078), The Science and Technology Research Program of Chongqing Municipal Education Commission (KJZD-M201800601), Sichuan Science and Technology Program (2021YFQ0053)

1 引言

智能驾驶的发展在很大程度上依赖于机器学习和人工智能技术,大多数智能驾驶任务,如安全/预警应用、辅助驾驶、信息娱乐应用和服务等都是计算密集型的^[1,2],由于传输时延、回程带宽有限和连接不稳定等原因,传统的集中式训练方案并不适用于这些任务。联邦学习是一种很有前途的分布式训练模型方法^[3,4]。智能驾驶中基于联邦学习的协同训练具有以下优点:保护数据隐私、提高模型性能以及提高系统成本效益。然而,当单个智能网联汽车(Intelligent Connected Vehicle, ICV)的计算和存储能力不足或出现故障时,联邦学习将在大规模协同训练方面表现较差,为了应对这个挑战,数字孪生技术提供了一种有效的解决方案。

数字孪生是一种将物理实体实时映射到数字空间的技术,可以实时捕获物理实体的动态状态信息^[5-7]。已有研究将联邦学习与数字孪生集成到各种网络系统中。文献^[8,9]利用数字孪生技术构建了所有工业设备或所有任务和路侧单元(Road Side Unit, RSU)的数字孪生体,以便从捕获设备特征或提供任务与RSU的匹配方案等方面辅助联邦学习,但并未考虑创建和维护数字孪生体所消耗的时延和能耗对联邦学习的影响。

因此,在训练时延、能耗与计算资源有限的智能驾驶场景中,考虑到维护ICV数字孪生体产生的额外能耗以及模型训练的收敛速度和质量,为了解决这些问题,本文研究了数字孪生辅助联邦学习中的边缘选择和资源分配联合优化,以克服ICV的资源有限性,提高联邦学习的训练效率。具体工作总结如下:

首先,提出了数字孪生辅助联邦学习机制,该机制支持ICV可动态选择在本地或利用其数字孪生体参与联邦学习。同时,在局部聚合阶段设计了质量加权聚合方法,通过量化本地模型的质量确定其参与局部聚合的权重,解决数字孪生体的映射偏差问题。其次,为提高数字孪生辅助联邦学习的训练效率,本文综合考虑数字孪生体产生的资源消耗和边缘服务器(Edge Server, ES)的可用资源,构建了数字孪生辅助联邦学习训练的计算和通信模型。通过分析边缘选择和联邦学习训练过程中的时间和能量消耗,建立以最小化联邦学习累积时延和能耗为目标的边缘选择和资源分配联合优化问题。最后,由于部分可观测的马尔可夫决策过程(Partially Observable Markov Decision Process, POMDP)模型中存在混合动作空间,本文提出基于多智能体参数化Q网络(Multi-agent Parametrized Deep Q-

Networks, MPDQN)的边缘选择和资源分配算法,在时变的驾驶环境中,探索近似最优的边缘选择和资源分配策略以实现联邦学习累积时延和能耗最小化。

2 系统模型

2.1 系统场景

智能驾驶场景下,数字孪生辅助联邦学习系统分为3层:终端层、边缘层和云层,如图1所示,终端层由具有有限计算和存储资源的ICV和具有监测能力的基础设施等实体组成。这些物理实体通过无线信道连接到边缘层中的基站,其中, N 辆ICV的集合可表示为 $\mathcal{N} = \{1, \dots, n, \dots, N\}$ 。ICV行驶过程中会通过传感器等收集到各种数据,这些本地数据可用来支持智能驾驶中各种任务的模型训练。边缘层包含 M 个具有ES的基站,其集合表示为 $\mathcal{M} = \{1, \dots, m, \dots, M\}$,ES $_m$ 下的ICV集合表示为 $\mathcal{N}_m = \{1, \dots, n, \dots, N_m\}$ 。本文假设一个ES可以创建和维护多个物理实体或多个物理子系统的数字孪生体。云层中包含具有参数服务器的中心云,可表示为 B ,参数服务器拥有更强大的计算、通信和存储能力,在该系统中主要用于局部模型的全局聚合。

根据文献^[10]可知,数字孪生体被建模为设备数字孪生体和业务数字孪生体两种。本文主要涉及设备数字孪生体,具体来说,ICV的数字孪生体可表示为: $DT_n = \{C_n, d_n, H_n(k), T_{n-m}\}$,其中, C_n 为ICV $_n$ 的行为模型; d_n 为ICV $_n$ 的基本数据(如ICV基本配置信息和驾驶员标签等); $H_n(t) = \{h_n^1(k), h_n^2(k), \dots, h_n^r(k)\}$ 为在时隙 k ICV $_n$ 在行驶过程中的

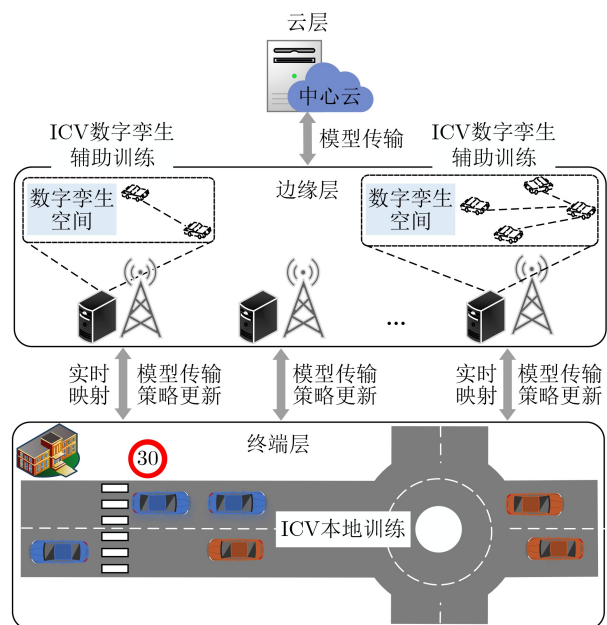


图1 数字孪生辅助联邦学习架构图

r 维运行状态, $H_n(k)$ 可以为加速、减速、滑行及制动等; \mathcal{T}_{n-m} 表示数字孪生体被维护的时间段, 如 $\mathcal{T}_{n-m} = (k_i, k_j)$ 表示ICV $_n$ 在 $k_i \sim k_j$ 时间段内选择在基站 m 中维护其数字孪生体。考虑动态的信道状态对ICV影响, 信道数字孪生体建模为 $DT_{\text{channel}}(t) = \{h_{n,m}(k), W_{n,m}(k) | \forall n \in \mathcal{N}_m, \forall m \in \mathcal{M}\}$, 其中 $h_{n,m}^{\text{DT}}(k)$ 与 $W_{n,m}(k)$ 分别表示ICV $_n$ 到ES $_m$ 之间的信道增益以及带宽。

2.2 数字孪生辅助联邦学习机制

本节从边缘选择模型和数字孪生辅助联邦学习的训练过程两个方面介绍数字孪生辅助联邦学习机制。

(1)边缘选择模型: 本文采用二进制变量 φ 表示ICV的边缘选择策略, 具体表示为: $\varphi_{n,m} \in \{0, 1\}$ 。 $\varphi_{n,m} = 0$ 表示ICV $_n$ 未选择在ES $_m$ 中维护其数字孪生体, 而是在本地完成模型训练, 并将模型参数上传给最近的ES进行局部聚合; $\varphi_{n,m} = 1$ 表示ICV $_n$ 选择在ES $_m$ 中维护其数字孪生体, 即选择利用数字孪生体在ES $_m$ 中进行模型训练。本文假设 $\varphi'_{n,m}$ 表示ICV $_n$ 是否关联到ES $_m$, 具体表示为

$$\varphi'_{n,m} = \begin{cases} 1, & \varphi_{n,m} = 0 \\ 0, & \varphi_{n,m} = 1 \end{cases} \quad (1)$$

本文假设每个ES最多能为 M_{max} 辆ICV提供数字孪生服务, 由此可得 $\sum_{n=1}^{N_m} \varphi_{n,m} \leq M_{\text{max}}$, 同样的, 每个ICV最多选择一个ES进行关联, 即 $\sum_{m=1}^M \varphi'_{n,m} \leq 1$ 。此外, 考虑到数字孪生的实时性, 本文假设用于维护ICV数字孪生体的传感器数据和模型仅在训练模型期间映射到服务器。

创建数字孪生体产生的时延和能耗是决定边缘选择的关键因素。本文将创建ICV $_n$ 的数字孪生体所需时延表示为 $T_{n,m}^{\text{DT}_1} = (\lambda_m |d_{n,m}|) / f_m$, 其中, λ_m 为ES处理一个字节数据所需的CPU周期数; $|d_{n,m}|$ 为ICV $_n$ 用于创建其数字孪生体的数据量大小; f_m 为ES的可用CPU频率。同时, 创建数字孪生体所产生的能耗表示为 $E_{n,m}^{\text{DT}_1} = \sigma_1 T_{n,m}^{\text{DT}_1}$, 其中, σ_1 为ES的计算能力。

(2)数字孪生辅助联邦学习的训练过程:

边缘选择和资源分配: ICV从ES中获取本轮全局更新中近似最优的边缘选择策略, 即确定在本轮全局更新中是否选择借助数字孪生辅助模型训练。另外, 需要维护ICV数字孪生体的ES获取本轮全局更新中近似最优的资源分配策略, 即为每个ICV数字孪生体分配多少计算资源进行模型训练和孪生体实时更新。

本地训练和数字孪生辅助训练: 每个ICV根据具体的边缘选择策略进行模型训练。本地训练是指

ICV利用本地数据集和自身计算资源完成模型训练。数字孪生辅助训练是指ICV数字孪生体在数字孪生空间中利用ES的计算资源完成模型训练。

本地模型上传: 完成本地训练的ICV及时将最新的模型参数上传给最近的ES。

局部聚合: ES需要聚合来自通信范围内的ICV本地模型和数字孪生空间中的本地模型。本文考虑到高质量的模型参数在聚合中占据更大的比重会提高模型精度, 也可以实现抵御恶意攻击。模型质量表示为

$$q_{n,m} = \frac{p_{n,m}}{\hat{f}_{n,m}} \quad (2)$$

其中, $\hat{f}_{n,m}$ 表示ES $_m$ 中ICV $_n$ 孪生体的CPU频率映射偏差; 若ICV $_n$ 并未选择在ES上利用数字孪生体进行模型训练, $\hat{f}_{n,m}$ 则定为常值; $p_{n,m}$ 是模型贡献值, 具体计算方式为: $p_{n,m} = (\omega_{n,m} - \bar{\omega}_m) / (\sum_{n=1}^{N_m} \omega_{n,m} - \bar{\omega}_m)$ 。其中, $\omega_{n,m}$ 表示ES $_m$ 中ICV $_n$ 的模型参数; $\bar{\omega}_m$ 表示ES $_m$ 中所有模型参数的平均值。

在局部聚合时, ES首先更新各ICV的模型质量值, 并根据下式将所有的本地模型聚合为局部模型 ω_m

$$\omega_m = \sum_{n=1}^{N_m} \frac{\omega_{n,m} q_{n,m} |D_{n,m}|}{|D_m|} \quad (3)$$

其中, $|D_{n,m}|$ 表示ES $_m$ 通信范围内的ICV $_n$ 的数据样本大小; $|D_m| = \sum_{n=1}^{N_m} |D_{n,m}|$ 表示ES $_m$ 范围内所有参与联邦学习的ICV的数据样本大小。

局部模型上传: 为减少ES与中心云的通信轮数, 本文选择在每隔 t' 个全局迭代轮次, 各ES将最新的局部模型参数上传给中心云。

全局聚合: 中心云在收到ES上传的局部模型参数后, 根据下式及时完成模型的全局聚合, 以得到最新的全局模型 ω

$$\omega = \sum_{m=1}^M \frac{\omega_m \alpha_m |D_m|}{(\alpha_m + \beta_m) |D|} \quad (4)$$

其中, α_m 是在ES $_m$ 中有积极贡献的模型数量; β_m 是ES $_m$ 中有恶意的模型数量; $|D| = \sum_{m=1}^M |D_m|$ 表示所有ICV的数据样本大小。

全局模型更新: 中心云将最新的全局模型下发给每个参与联邦训练的ICV和ES。ICV和ES再重复下一次迭代, 直到整个训练过程收敛。

边缘选择和资源分配策略更新: 本文采用MP-DQN算法探索每轮全局迭代中近似最优的边缘选择和资源分配策略, 最小化数字孪生辅助联邦学习

机制中的累积训练时延和能耗。MPDQN算法通过不断与环境交互，为下一轮全局更新提供近似最优的边缘选择和资源分配策略。

2.3 数字孪生辅助联邦学习的计算模型

计算模型包括ICV使用自身实时数据进行模型训练的本地计算、ES使用数字孪生空间中ICV数字孪生体进行模型训练的边缘计算、ES聚合本地模型参数的局部聚合计算和中心云聚合局部模型的全局聚合计算。假设训练一个字节数据所需CPU周期数为 λ_1 ，聚合一个字节数据所需CPU周期数为 λ_2 ，在时隙 k （在后文表示为第 k 轮）分配给ICV n 的数字孪生体训练的CPU周期频率为 $f_{n,m}(k)$ 。

ICV n 本地训练的计算时间表示为

$$T_{n,m}^{\text{com1}}(k) = \frac{\lambda_1 \varphi'_{n,m} |D_{n,m}(k)|}{f_n} \quad (5)$$

其中， $|D_{n,m}(k)|$ 为在时隙 k 即第 k 轮全局迭代中ICV n 参与训练的数据大小； f_n 为ICV n 本地训练的CPU周期频率。同理，ICV n 的数字孪生体在ES m 中的训练时间表示为

$$T_{n,m}^{\text{com2}}(k) = \frac{\lambda_1 \varphi_{n,m} D_{n,m}(k)}{f_{n,m}(k)} \quad (6)$$

ES聚合本地模型参数的聚合时间表示为

$$T_m^{\text{com}}(k) = \frac{\lambda_2 \sum_{n=1}^{N_m} |\omega_{n,m}|}{f_m(k)} \quad (7)$$

其中， $|\omega_{n,m}|$ 表示ES m 通信范围内的ICV n 的模型大小。 M 个ES总聚合时间为 $T_M^{\text{com}}(k) = \max\{T_m^{\text{com}}(k)\}$ 。

中心云聚合ES所上传的局部模型的计算时间为

$$T_{B_d}^{\text{com}}(k) = \frac{\lambda_1 \sum_{m=1}^M |\omega_m|}{f_B} \quad (8)$$

其中， $|\omega_m|$ 为ES m 上传的局部模型大小； f_B 表示中心云 B 聚合模型的CPU周期频率。

系统的计算能耗分为5个部分，包括ICV本地训练的计算能耗 $E_{n,m}^{\text{com1}}(k) = \sigma_2 T_{n,m}^{\text{com1}}(k)$ 、ES使用数字空间中ICV数字孪生体训练模型的计算能耗 $E_{n,m}^{\text{com2}}(k) = \sigma_1 T_{n,m}^{\text{com2}}(k)$ 、ES聚合本地模型的聚合能耗 $E_m^{\text{com}}(k) = \sigma_3 T_m^{\text{com}}(k)$ 、ES维护ICV数字孪生体的能耗 $E_m^{\text{DT}_2}(k) = \sum_{n=1}^{N_m} \varphi_{n,m} \sigma_4 |d_{n,m}(k)|$ 以及中心云聚合局部模型的聚合能耗 $E_B^{\text{com}}(k) = \sigma_5 T_B^{\text{com}}(k)$ ，其中， σ_2 表示ICV的计算能力； σ_3 表示ES的聚合能力； σ_4 表示ES在数字空间中维持单个数字孪生体的能力； σ_5 表示中心云的聚合能力。

2.4 数字孪生辅助联邦学习的通信模型

在数字孪生辅助的联邦学习中，ICV、ES和中

心云之间需要频繁传输模型参数完成模型更新，以实现预设精度或模型收敛。从ICV n 到ES m 以及ES m 到中心云 B 的数据速率分别为

$$r_{n,m}(k) = W_{n,m} \log_2(1 + \gamma_{n,m}(k)) \quad (9)$$

$$r_{k,B}(k) = W_m \log_2(1 + \gamma_{m,B}(k)) \quad (10)$$

其中， $W_{n,m}$ 表示ICV n 到ES m 的信道带宽； $\gamma_{n,m}(k) = h_{n,m}(k) P_n^{\text{tran}} / N_0$ 表示ICV n 和ES m 之间的信道状态， $h_{n,m}(k)$ 表示ICV n 到ES m 之间的信道增益； P_n^{tran} 表示ICV n 的传输功率； N_0 表示噪声功率； W_m 表示ES m 的信道带宽； $\gamma_{m,B}(k)$ 表示ES m 和中心云 B 之间的信道状态。

在通信模型中，传输时间包括ICV到ES、ES到中心云以及中心云到ES和ICV 3个部分，中心云下发全局模型参数所消耗的时间忽略不计。因此，从ICV n 到ES m 以及ES m 到中心云 B 的传输时间分别为

$$T_{n,m}^{\text{tran}}(k) = \frac{\varphi'_{n,m} |\omega_n|}{r_{n,m}(k)} \quad (11)$$

$$T_{m,B_d}^{\text{tran}}(k) = \frac{|\omega_m|}{r_{m,B}(k)} \quad (12)$$

传输能耗同样包括ICV到ES $E_{n,m}^{\text{tran}}(k) = T_{n,m}^{\text{tran}}(k) P_{n,m}^{\text{tran}}$ 、ES到中心云 $E_{k,B}^{\text{tran}}(k) = T_{k,B}^{\text{tran}}(k) P_k^{\text{tran}}$ 以及中心云到ES和ICV 3个部分。其中， $P_{n,m}^{\text{tran}}$ 为ICV n 的传输功率； P_m^{tran} 为ES m 的传输功率。中心云到ES和ICV这部分的能耗忽略不计。

2.5 优化目标

边缘选择和资源分配联合优化的目标是通过为ICV选择模型训练方式和为ICV数字孪生体分配计算资源，使得数字孪生辅助联邦学习训练的累积时延和能耗最小化。

对于ICV n ，综合考虑数字孪生辅助联邦学习训练过程中的计算和通信时延。在第 t 次全局迭代中，ICV n 产生的总时延表示为

$$T_{n,m}(k) = T_{n,m}^{\text{DT}_1}(k) + T_{n,m}^{\text{com1}}(t) + T_{n,m}^{\text{com2}}(k) + T_{n,m}^{\text{tran}}(k) \quad (13)$$

考虑到同步聚合会耗费大量时间等待所有ICV上传其本地模型，本文采用半异步聚合方式，即设定ICV模型上传的截止时间 t_{\max} ，当 $T_{n,m}(k) > t_{\max}$ 时，ES选择不再等待该ICV，直接聚合已收到的本地模型参数。所有ICV的总时延由最慢完成本地训练的ICV决定，可表示为 $T_{N_m}(k) = \max_{n \in \mathcal{N}_m} \{T_{n,m}(k)\}$ 。

ES m 的总时延表示为 $T_m(k) = \max\{T_{N_m}(k), t_{\max}\} + T_m^{\text{com}}(k) + T_{m,B}^{\text{tran}}(k)$ 。在数字孪生辅助联邦学

习的一轮全局更新中，系统产生的总时延表示为：

$$T_{\text{all}}(k) = \max_{m \in \mathcal{M}} \{T_m(k)\} + T_B^{\text{com}}(k)。$$

ICV_{*n*}和ES_{*m*}产生的总能耗分别为

$$E_{n,m}(k) = \varphi'_{n,m} [E_{n,m}^{\text{com}1}(k) + E_{n,m}^{\text{tran}}(k)] \quad (14)$$

$$E_m(k) = E_{m,B}^{\text{tran}}(k) + \varphi_{n,m} E_m^{\text{com}}(k) + E_m^{\text{DT}1}(k) + E_m^{\text{DT}2}(k) \quad (15)$$

其中， $E_m^{\text{DT}1}(k)$ 为ES_{*m*}中创建ICV数字孪生体的总能耗，具体表示为： $E_m^{\text{DT}1}(k) = \sum_{n=1}^{N_m} [\varphi_{n,m} E_{n,m}^{\text{DT}1}(k)]$ 。

在数字孪生辅助联邦学习的一轮全局更新中，系统所产生的总能耗表示为 $E_{\text{all}}(k) = \sum_{m=1}^M [E_m(k) + \sum_{n=1}^{N_m} E_{n,m}(k)] + E_B^{\text{com}}(k)$ 。

根据以上对训练时延和能耗的分析，该优化目标可以表示为

$$\begin{aligned} & \min_{\varphi, f} \left\{ \sum_{k=1}^K [\mu_1 T_{\text{all}}(k) + \mu_2 E_{\text{all}}(k)] \right\} \\ & \text{s.t. } \left. \begin{aligned} & \text{C1: } F(\omega^K) - F(\omega^*) \leq \epsilon \\ & \text{C2: } \varphi_{n,m} \in \{0, 1\}, \forall n \in \mathcal{N}_m, \forall m \in \mathcal{M} \\ & \text{C3: } \sum_{m \in \mathcal{M}} \varphi_{n,m} \leq 1, \forall n \in \mathcal{N} \\ & \text{C4: } \sum_{n \in \mathcal{N}_m} \varphi_{n,m} \leq M_{\text{max}}, \forall m \in \mathcal{M} \\ & \text{C5: } f_{\text{min}} \leq f_{n,m} \leq f_{\text{max}}, \forall n \in \mathcal{N}_m, \forall m \in \mathcal{M} \end{aligned} \right\} \quad (16) \end{aligned}$$

其中， φ 表示ICV的边缘选择结果集合， f 表示ES中计算资源分配结果集合； μ_1 和 μ_2 分别为时延和能耗的系数。上述约束条件中，C1约束保证ML模型在 K 轮全局更新后，收敛至目标损失 $F(\omega^*)$ ，其中， ϵ 为收敛阈值，表示模型训练的目标精度；C2约束保证ICV_{*n*}只能选择在本地训练或者在ES_{*m*}中利用其数字孪生体辅助训练；C3约束保证ICV_{*n*}只能在一个ES中创建其数字孪生体；C4约束保证在单个ES中维护ICV数字孪生体的总数量不超过 M_{max} ；C5约束保证ES_{*m*}为ICV_{*n*}的数字孪生体分配到的CPU频率大于最小值 f_{min} 且小于最大值 f_{max} 。

3 基于MPDQN的边缘选择和资源分配算法

在智能驾驶场景中，ICV和ES的资源状态和任务列表在数字孪生辅助联邦学习的不同迭代中动态变化，且当ICV和ES数量增加时，上述优化问题的规模迅速增大。同时，该优化问题包含离散和连续的优化变量，且优化变量之间相互关联。这些都增加了问题的复杂性，使得传统的优化方法难以求出最优解。本文采用基于MPDQN的边缘选择和资源分配算法，该算法分为集中训练和独立执行两个阶段。

3.1 POMDP

为了优化数字孪生辅助联邦学习的训练过程，本文首先将联合优化问题转化为适用于DRL求解的POMDP问题。POMDP模型由以下元素 $\langle \mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R} \rangle$ 表示，在这个4元组中， $\mathcal{S} = \{\mathcal{S}_1, \dots, \mathcal{S}_m, \dots, \mathcal{S}_M\}$ 表示全局状态空间，其中 \mathcal{S}_m 为智能体 m 可观测到的部分观测信息； $\mathcal{A} = \{\mathcal{A}_1 \times \dots \times \mathcal{A}_m \times \dots \times \mathcal{A}_M\}$ 表示 M 个分别被部署于ES中智能体的联合动作空间。下面详细介绍智能体 m 的4元组 $\langle \mathcal{S}_m, \mathcal{A}_m, \mathcal{P}_m, \mathcal{R} \rangle$ ：

(1)状态空间：状态空间 \mathcal{S}_m 指智能体 m 观测通过在数字孪生体的辅助下观察到的ICV和ES_{*m*}的资源情况和状态信息等。具体包含：ES_{*m*}的可用计算资源集合 $\varsigma_m(k) = \{\varsigma_{m,1}(k), \dots, \varsigma_{m,N_m}(k), \varsigma^m(k)\}$ ，其中， $\varsigma_{m,N_m}(k)$ 表示ES_{*m*}覆盖范围内ICV_{*N_m*}的可用计算资源， $\varsigma^m(k)$ 表示ES_{*m*}的可用计算资源；ES_{*m*}的通信资源集合 $\gamma_m(k) = \{\gamma_{m,1}(k), \dots, \gamma_{m,N_m}(k), \gamma^m(k)\}$ 中的 $\gamma_{m,N_m}(k)$ 表示ES_{*m*}覆盖范围内ICV_{*N_m*}的通信资源， $H_m(k) = \{H_{m,1}(k), \dots, H_{m,N_m}(k)\}$ 表示ES_{*m*}覆盖范围内ICV_{*N_m*}的行为状态； $\gamma^m(k)$ 表示ES_{*m*}的通信资源；联邦学习全局模型当前的训练损失函数 $F(\omega^k)$ ；联邦学习在当前时刻的训练时延 $T_{\text{all}}(k)$ 和能耗 $E_{\text{all}}(k)$ 。在决策时隙 k ，系统状态 $s_{m,k} \in \mathcal{S}_m$ 可以表示为 $s_{m,k} = (\varsigma_m(k), \gamma_m(k), F(\omega^k), T_{\text{all}}(k), E_{\text{all}}(k))$ 。

(2)动作空间：智能体 m 的动作空间可以定义为 $\mathcal{A}_m = \{a_{m,k}\}$ ，其中 $a_{m,k} = \{\varphi_m(k), f_m(k)\}$ 。 $\varphi_m(k)$ 包含在ES_{*m*}下 N_m 辆ICV的边缘选择策略，其中 $\varphi_{n,m}(k) \in \{0, 1\}$ 。 $f_m(k)$ 包含在ES_{*m*}中为 N_m 个ICV数字孪生体分配CPU频率的资源分配策略。其中 $f_{\text{min}} \leq f_{n,m} \leq f_{\text{max}}$ ， f_{min} 为ES能分配给ICV数字孪生体 n 的最小CPU频率， f_{max} 为ES能分配给ICV数字孪生体 n 的最大CPU频率。

(3)状态转移概率：在POMDP中，状态转移概率被定义为一个条件概率分布，它给出了在给定当前状态 $s_{m,k}$ 和动作 $a_{m,k}$ 的情况下转移到新状态 $s_{m,k+1}$ 的概率，记为 $P_m(s_{m,k+1} | s_{m,k}, a_{m,k})$ 。

(4)奖励函数：本文的目标是在保证模型精度的同时，最小化数字孪生辅助联邦学习训练过程中的累积时延和能耗。因此，瞬时奖励函数应同时关注联邦学习训练时延和能耗以及全局模型损失。瞬时奖励函数 $r_{m,k}$ 具体内容如下所示

$$r_{m,k} = \vartheta_1 (\mu_1 T_{\text{all}}(k) + \mu_2 E_{\text{all}}(k)) + \vartheta_2 \frac{F(\omega^k) - F(\omega^*)}{F(\omega^*)} \quad (17)$$

其中， ϑ_1 和 ϑ_2 分别是联邦训练时延和能耗以及模型损失的权重， $\vartheta_1, \vartheta_2 \in [-1, 0]$ 。

3.2 基于MPDQN 的边缘选择和资源分配算法

本文的优化变量涉及边缘选择和资源分配，则动作空间中既包含离散动作 $\varphi_{n,m}(k) \in \{0, 1\}$ ，又包含连续动作 $f_{n,m} \in [f_{\min}, f_{\max}]$ 。本文引用PDQN算法求解混合动作空间的优化问题，PDQN算法结合了DQN和DDPG的优点，使得它在处理边缘选择和资源分配联合优化中的混合动作空间问题有较好的效果。

如图2所示，本文使用 $Q_m(s_m, a_m) = Q_m(s_m, \varphi_m, f_m)$ 表示智能体 m 的动作值函数，其中，状态 $s_m \in \mathcal{S}_m$ ， $\varphi_m = \{\varphi_{n,m} | \forall n \in \mathcal{N}_m\}$ 为离散动作的集合， f_m 为连续动作的集合，此时贝尔曼方程变为

$$Q_m(s_{m,k}, \varphi_m(k), f_m(k)) = \mathbb{E}_{\pi} [r_{m,k} + \ell \max_{\varphi_m, f_m} Q_m(s_{m,k+1}, \varphi_m(k+1), f_m(k+1)) | s_{m,k} = s_m, a_{m,k} = (\varphi_m(k), f_m(k))] \quad (18)$$

根据文献[11]，当 $f_m^*(k) = \arg \sup_{f_m} Q_m(s_{m,k+1}, \varphi_m(k+1), f_m(k+1))$ 取最大值时，可获得最优的连续动作 $f_m^*(k)$ 。当 Q 函数固定时，对于任意的系统状态和离散动作，可将 $\arg \sup_{f_m} Q_m(s_m, \varphi_m, f_m)$ 视为系统状态到连续动作的映射函数： $f_m^{Q_m} : \mathcal{S}_m \rightarrow [f_{\min}, f_{\max}]$ 。为此贝尔曼方程可改写为

$$Q_m(s_{m,k}, \varphi_m(k), f_m(k)) = \mathbb{E}_{\pi} [r_{m,k} + \ell \max_{\varphi_k} Q_m(s_{m,k+1}, \varphi_m(k+1), f_m^{Q_m}(k+1)) | s_{m,k} = s_m] \quad (19)$$

本文选择使用DNN的 $Q_m(s_m, \varphi_m, f_m; \varpi_m)$ 去

近似动作值函数 $Q_m(s_m, \varphi_m, f_m)$ ，其中， ϖ_m 为网络权重。根据离散动作值函数的数值大小选择出最优的离散动作，即最大的离散动作值函数对应的离散动作为最优的离散动作。因此，最优的离散动作可表示为 $\varphi_m^*(t) = \arg \max_{\varphi_m} Q_m(s_{m,k}, \varphi_m(k), f_m^*(k))$ 。

为保证PDQN算法训练的稳定性，本文在PDQN中引入目标 Q 网络(Target Q Network, TQN)。TQN与 Q 网络的神经网络结构相同，且TQN的网络参数是在一定时间间隔内从 Q 网络中复制得到。因此，TQN的目标值为： $y_m(k) = r_{m,k} + \ell \max_{\varphi_m} Q_m(s_{m,k}, \varphi_m(k), f_m(k); \varpi_{m,c}, \varpi'_{m,d})$ 。其中， $\varpi'_{m,d}$ 为智能体 m 中TQN的网络参数，具体为不同时刻 Q 网络中网络参数 $\varpi_{m,d}$ 的副本。

PDQN算法中 Q 网络和确定性策略网络的损失函数分别可表示为

$$L(\varpi_{m,d}) = \mathbb{E}_{\pi} [y_m(k) - Q_m(s_{m,k}, \varphi_m(k), f_m(k); \varpi_{m,c}, \varpi_{m,d})]^2 \quad (20)$$

$$L(\varpi_{m,c}) = - \sum_{\varphi_m} Q_m(s_{m,k}, \varphi_m(k), f_m(k); \varpi_{m,c}) \quad (21)$$

Q 网络和确定性策略网络的网络参数通过梯度下降法进行更新，具体表示为

$$\varpi_{m,d}(k+1) = \lambda_s \cdot \left(\varpi_{m,d}(k) - \lambda_d \cdot \frac{\partial L(\varpi_{m,d})}{\partial \varpi_{m,d}} \Big|_{\varpi_{m,d} = \varpi_{m,d}(k)} \right) + \lambda_{me} \cdot \varpi_{d,me}(k) \quad (22)$$

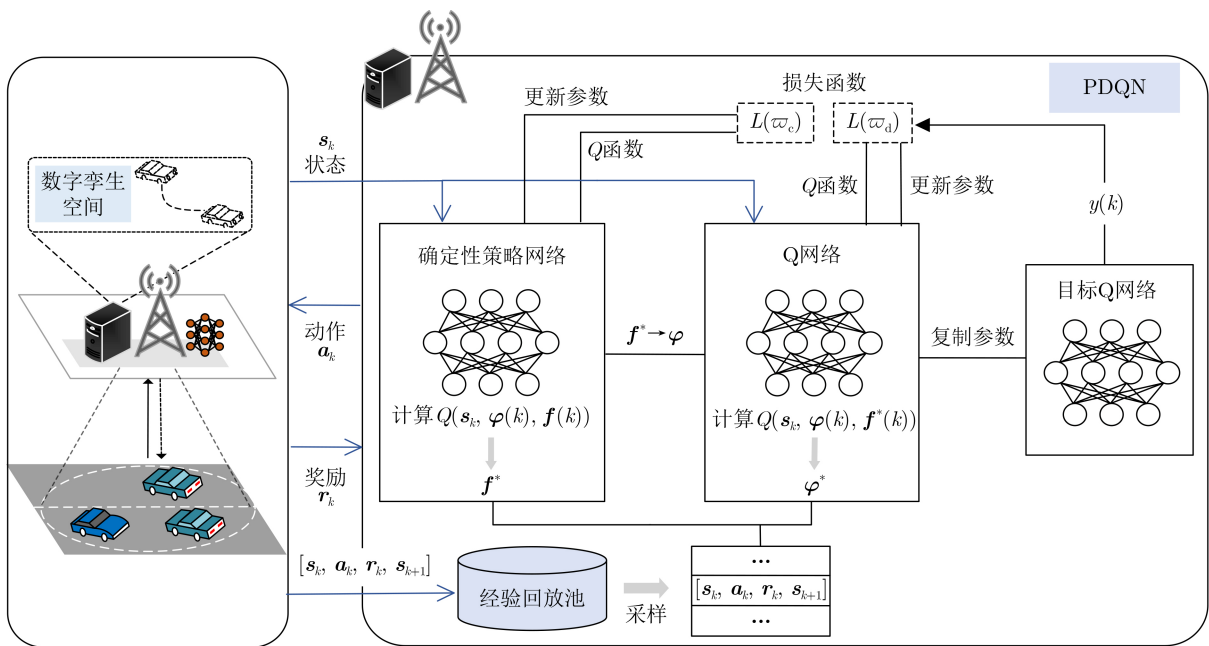


图2 单个ES中的PDQN框架

$$\begin{aligned} \varpi_{m,c}(k+1) = & \lambda_s \cdot \left(\varpi_{m,c}(k) - \lambda_c \right. \\ & \left. \cdot \frac{\partial L(\varpi_{m,c})}{\partial \varpi_{m,c}} \Big|_{\varpi_{m,c} = \varpi_{m,c}(k)} \right) \\ & + \lambda_{me} \cdot \varpi_{c,me}(k) \end{aligned} \quad (23)$$

其中, λ_d 和 λ_c 分别为Q网络和确定性策略网络的学习率; $\varpi_{d,me}$ 和 $\varpi_{c,me}$ 分别为融合网络更新的两个网络参数, 其具体内容在后文详细给出; λ_s 和 λ_{me} 分别为智能体的本地参数和融合参数的权重。

本文采用多智能体“集中训练, 独立执行”的方式完成协同学习, 以实现最优的边缘选择和资源分配策略。在集中训练阶段, 中心云中的融合网络^[12]将融合所有智能体, 对各个智能体的Q函数进行非线性组合。融合网络以全局状态 S 和各智能体的 Q_m 函数作为输入, 得到全局Q函数 $Q_{me} = f_{me}(Q_{gl})$ 。其中, $f_{me}(\cdot)$ 表示融合网络; $Q_{gl} = \{Q_1, \dots, Q_m, \dots, Q_M\}$ 为 M 个智能体的 Q_m 函数的非线性组合。融合网络通过最小化全局损失函数 $l(\varpi)$ 来更新其参数

$\varpi = \{\varpi_{1,d}, \dots, \varpi_{m,d}, \dots, \varpi_{M,d}, \varpi_{me}\}$, 其中, ϖ_{me} 为融合网络的网络参数。融合网络的损失函数表示为 $l(\varpi) = \mathbb{E}[y_{me} - Q_{me}(s_k, \varphi_k, f_k)]^2$, 其中, $y_{me} = r + \ell \max_{\varphi} \varphi Q_{me}(s_k, \varphi_k, f_k(\varpi'_d))$; s_k 表示全局状态。智能体 m 的所有离散动作 Q_m 值求和得到的 \tilde{Q}_m , 将所有智能体的 \tilde{Q} 值作为融合网络的输入得到 Q_{me}^s , 即 $Q_{me}^s = f_{me}(s, \tilde{Q}_1, \dots, \tilde{Q}_m, \dots, \tilde{Q}_M; \varpi)$ 。

当参数 ϖ 固定时, 通过最大化 Q_{me}^s 来更新所有智能体的确定性策略网络参数, 其梯度可表示为

$$\begin{aligned} \nabla_{\varpi_{m,c}} l(\varpi_{m,c}) = & \nabla_{\tilde{Q}_m} f_{me}(s, \tilde{Q}_1, \dots, \tilde{Q}_m, \dots, \tilde{Q}_M; \varpi) \\ & \cdot \nabla_{\varpi_{m,c}} \tilde{Q}_m(s_k, \varphi_k, f_k; \varpi) \Big|_{f_{m,k} = f_m(\cdot)} \end{aligned} \quad (24)$$

数字孪生辅助联邦学习机制中, 基于MPDQN的边缘选择和资源分配算法如算法1所示。

4 仿真结果与分析

智能驾驶场景中包含大量计算密集型任务, 在仿真实验中可以通过在ICV本地训练图像数据集,

算法1 基于MPDQN的边缘选择和资源分配算法

输入: 学习率(λ_d, λ_c), 学习回合数 N_{max} , 概率分布 ψ , 探索概率 ε , 小批量大小 B , 采样数据的学习回合数量 N_{sam}

输出: 边缘选择和资源分配策略

- (1) 初始化网络参数(θ_d, θ_c)和经验回放池
- (2) **for** $i = 1, 2, \dots, N_{max}$ **do**
- (3) 收到初始状态 $s_1 = \{s_{m,1}\}_{\forall m \in \mathcal{M}}$
- (4) **for** 数字孪生辅助联邦学习全局迭代 $k \in \mathcal{K}$ **do**
- (5) **for** 智能体 $m \in \mathcal{M}$ **do**
- (6) 根据式(22)计算连续动作参数 $f_m(k)$
根据 ε 贪婪策略选择动作 $a_{m,k} = \{\varphi_m(k), f_m(k)\}$:
- (7)
$$a_{m,k} = \begin{cases} \text{分布}\psi\text{的样本}, & \varepsilon \\ (\varphi_m(k), f_m(k)), & \varphi_m(k) = \operatorname{argmax}_{\varphi} Q(s_{m,k}, \varphi_m(k), f_m^*(k)), 1 - \varepsilon \end{cases}$$
- (9) 执行动作 $a_{m,k}$, 获得瞬时奖励 $r_{m,k}$ 和下一个状态 $s_{m,k+1}$
- (10) 将元组 $[s_{m,k}, a_{m,k}, r_{m,k}, s_{m,k+1}]$ 存入经验回放池 D_m
- (11) 经验回放池 D_m 中采样一组小批量 B 的数据样本
- (12) 根据式(19)更新TQN的目标函数 $y_m(k)$
- (13) 根据式(20)和式(21)分别计算损失函数 $L(\varpi_{m,d})$ 和 $L(\varpi_{m,c})$
- (15) 根据式(22)和式(23)更新网络参数 $\varpi_{m,d}(k+1)$ 和 $\varpi_{m,c}(k+1)$
- (17) **if** ($i > N_{sam}$) **then**
- (18) 从经验回放池 D 中采样一组小批量 B 的数据样本
- (19) 更新参数 $\varpi_{d,me}(k+1) \leftarrow \lambda'_{me} \nabla_{\varpi_{d,me}} l(\varpi)$ 和 $\varpi_{c,me}(k+1) \leftarrow \lambda'_{me} \nabla l(\varpi_c)$
- (21) 融合网络下发最新的参数至各个智能体
- (22) **end if**
- (23) **end for**
- (24) **end for**
- (25) **end for**

以模拟真实的智能驾驶应用,如交通标志检测、交通信号灯检测和行人识别等。本文采用基于Python3.7的仿真平台和Pytorch模块验证所提算法的性能。另外,实验所需的图像数据集选用开源数据集——服装数据集(Fashion MNIST, FMNIST),该数据集包含70 000张不同商品的正面图片(60 000个训练示例和10 000个测试示例)^[13]。为模拟智能驾驶场景,本文假设系统中有1个中心云,5个ES和50辆ICV,ICV随机分布在ES覆盖区域内。其他仿真参数参考文献^[14]和文献^[15]设置。

(1)数字孪生的有效性:这部分仿真实验对比了传统联邦学习和数字孪生辅助联邦学习的训练性能,以证明数字孪生技术的有效性。在仿真实验中,用部分节点中途退出本地训练,以模拟资源不

足或出现故障的ICV。在传统联邦学习中选择直接舍弃这部分节点。图3和图4分别为不同联邦学习机制的训练损失和训练精度对比。两种联邦学习方案分别进行了600次迭代更新以实现模型收敛,在第200次全局迭代左右,两种方案的训练损失都趋于稳定,但传统联邦学习最终收敛的损失值略高于数字孪生辅助的联邦学习。同样的,传统联邦学习训练得到的全局模型精度也低于数字孪生辅助的联邦学习。这是因为数字孪生技术辅助计算资源不足的ICV继续参与联邦学习,即有效利用了各个ICV的本地数据,从而提高全局模型精度。

(2)边缘选择策略:这部分仿真实验用于验证边缘选择对联邦学习训练性能的影响。基准方案包括随机选择(Random Selection, RS)和就近选择(Nearby Selection, NS)。RS是指ICV随机选择是否在某个ES中维护其数字孪生体,然后使用PDQN算法中的确定性深度策略网络求解ES的资源分配问题。NS是指需要维护数字孪生体的ICV根据距离优先的原则选择在某个ES中维护其数字孪生体,资源分配问题的处理方法同随机选择。图5和图6分别为不同边缘选择策略的训练精度以及训练时延和能耗对比,从这3个方面对比不同的边缘选择策略可以看出本文方案收敛更快,模型精度更高,训练时延和能耗更低,NS次之,RS最差。本文方案的

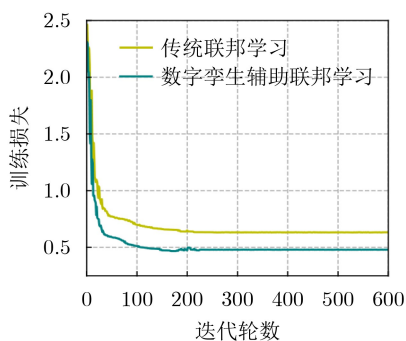


图3 不同联邦学习机制的训练损失对比

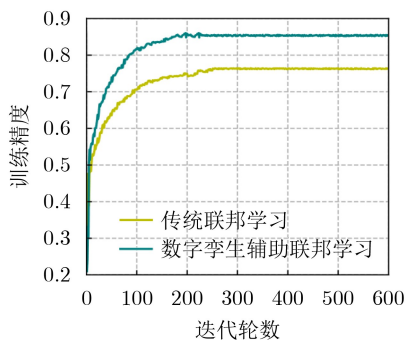


图4 不同联邦学习机制的训练精度对比

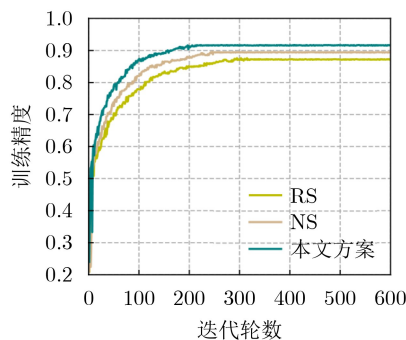
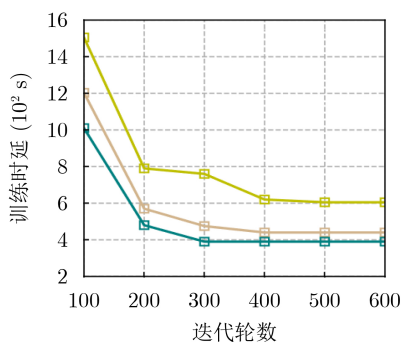
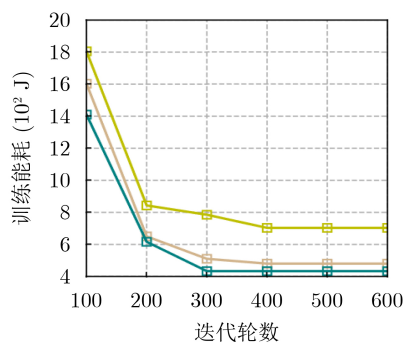


图5 不同边缘选择策略的训练精度对比



(a) 训练时延对比



(b) 训练能耗对比

图6 不同边缘选择策略的时延和能耗对比

性能略优于NS, 这是因为ICV与ES的距离是影响选择结果的关键因素。但本文方案联合考虑了距离和ES资源可用率等多种因素, 这使得所做出的策略更适用于变化的网络环境。而RS的不确定性大且不稳定。本文方案、RS和NS 3种方案最终收敛的训练精度约为91%, 87%和89%。

(3)资源分配策略: 这部分仿真实验用于验证资源分配对联邦学习训练性能的影响。基准方案包括随机分配(Random Allocation, RA)和平均分配(Average Allocation, AA)。RA是指ES根据随机比例分配计算资源, AA是指ES按照均分原则分配计算资源。图7和图8分别为训练精度以及训练时延和能耗对比, 从这3副图可以看出本文方案最优, AA次之, RA最差。其中, RA不确定性较大, 导致其收敛速度较慢且会产生更多的训练时延和能

耗。AA基于平均的原则分配资源造成资源分配不均, 从而导致更多的训练时延和能耗。本文方案、RA和AA 3种方案最终收敛的训练精度约为91%, 78%和87%。

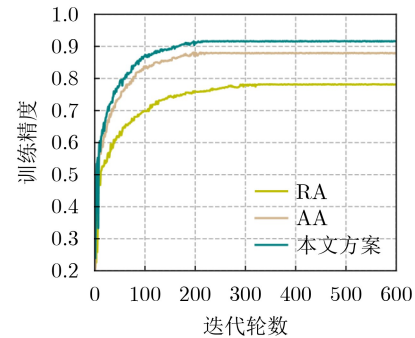


图7 不同资源分配策略的训练精度对比

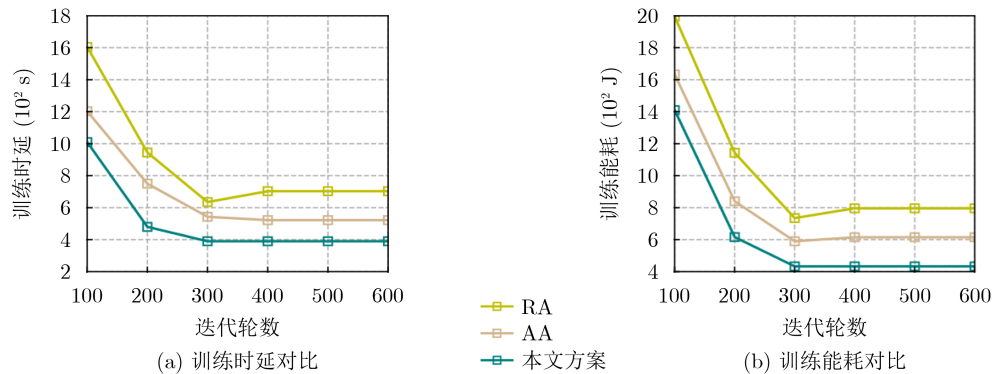


图8 不同资源分配策略的时延和能耗对比

5 结束语

针对基于联邦学习的智能驾驶中, ICV资源受限或突发故障会导致联邦学习训练精度低、时延长和能耗高等问题, 提出了数字孪生辅助联邦学习中的边缘选择和资源分配优化算法。首先, 数字孪生辅助联邦学习机制支持ICV在本地参与联邦学习或利用其数字孪生体参与联邦学习。其次, 通过构建数字孪生辅助联邦学习训练的计算和通信模型, 建立以最小化联邦学习累积时延和能耗为目标的联合优化问题, 并将其转化为POMDP。最后, 提出基于MPDQN的边缘选择和资源分配算法, 探索策略以实现联邦学习累积时延和能耗最小化。仿真结果表明, 所提算法显著减少联邦学习的累积训练时延和能耗, 且不降低模型精度。

参考文献

- [1] BOUKERCHE A and DE GRANDE R E. Vehicular cloud computing: Architectures, applications, and mobility[J]. *Computer Networks*, 2018, 135: 171–189. doi: 10.1016/j.comnet.2018.01.004.
- [2] ARENA F and PAU G. An overview of vehicular communications[J]. *Future Internet*, 2019, 11(2): 27. doi: 10.3390/fi11020027.
- [3] BENNIS M. Federated learning and control at the wireless network edge[J]. *GetMobile: Mobile Computing and Communications*, 2021, 24(3): 9–13. doi: 10.1145/3447853.3447857.
- [4] CHEN Mingzhe, POOR H V, SAAD W, et al. Convergence time minimization of federated learning over wireless networks[C]. ICC 2020–2020 IEEE International Conference on Communications (ICC), Dublin, Ireland, 2020: 1–6. doi: 10.1109/ICC40277.2020.9148815.
- [5] WU Yiwen, ZHANG Ke, and ZHANG Yan. Digital twin networks: a survey[J]. *IEEE Internet of Things Journal*, 2021, 8(18): 13789–13804. doi: 10.1109/JIOT.2021.3079510.
- [6] GRIEVES M and VICKERS J. Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems[M]. KAHLEN F J, FLUMERFELT S, and ALVES A. *Transdisciplinary Perspectives on Complex Systems*:

- New Findings and Approaches. Cham, Germany: Springer, 2017: 85–113. doi: [10.1007/978-3-319-38756-7_4](https://doi.org/10.1007/978-3-319-38756-7_4).
- [7] DAI Yueyue, GUAN Yongliang, LEUNG K K, *et al.* Reconfigurable intelligent surface for low-latency edge computing in 6G[J]. *IEEE Wireless Communications*, 2021, 28(6): 72–79. doi: [10.1109/MWC.001.2100229](https://doi.org/10.1109/MWC.001.2100229).
- [8] SUN Wen, LEI Shiyu, WANG Lu, *et al.* Adaptive federated learning and digital twin for industrial internet of things[J]. *IEEE Transactions on Industrial Informatics*, 2021, 17(8): 5605–5614. doi: [10.1109/TII.2020.3034674](https://doi.org/10.1109/TII.2020.3034674).
- [9] HUI Yilong, ZHAO Gaosheng, LI Chengle, *et al.* Digital twins enabled on-demand matching for multi-task federated learning in HetVNs[J]. *IEEE Transactions on Vehicular Technology*, 2023, 72(2): 2352–2364. doi: [10.1109/TVT.2022.3211005](https://doi.org/10.1109/TVT.2022.3211005).
- [10] LU Yunlong, MAHARJAN S, and ZHANG Yan. Adaptive edge association for wireless digital twin networks in 6G[J]. *IEEE Internet of Things Journal*, 2021, 8(22): 16219–16230. doi: [10.1109/JIOT.2021.3098508](https://doi.org/10.1109/JIOT.2021.3098508).
- [11] XIONG Jiechao, WANG Qing, YANG Zhuoran, *et al.* Parametrized deep Q-networks learning: Reinforcement learning with discrete-continuous hybrid action space[J]. arXiv: 1810.06394, 2018. doi: [10.48550/arXiv.1810.06394](https://doi.org/10.48550/arXiv.1810.06394).
- [12] YIN Sixing and YU F R. Resource allocation and trajectory design in UAV-aided cellular networks based on multiagent reinforcement learning[J]. *IEEE Internet of Things Journal*, 2022, 9(4): 2933–2943. doi: [10.1109/JIOT.2021.3094651](https://doi.org/10.1109/JIOT.2021.3094651).
- [13] XIAO Han, RASUL K, and VOLLGRAF R. Fashion-MNIST: a novel image dataset for benchmarking machine learning algorithms[J]. arXiv: 1708.07747, 2017. doi: [10.48550/arXiv.1708.07747](https://doi.org/10.48550/arXiv.1708.07747).
- [14] YU Xiangbin, XU Weiye, LEUNG S H, *et al.* Power allocation for energy efficient optimization of distributed MIMO system with beamforming[J]. *IEEE Transactions on Vehicular Technology*, 2019, 68(9): 8966–8981. doi: [10.1109/TVT.2019.2931291](https://doi.org/10.1109/TVT.2019.2931291).
- [15] ZHANG Jiexiang, LIU Yiming, QIN Xiaoqi, *et al.* Energy-efficient federated learning framework for digital twin-enabled industrial internet of things[C]. The IEEE 32nd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), Helsinki, Finland, 2021: 1160–1166. doi: [10.1109/PIMRC50174.2021.9569716](https://doi.org/10.1109/PIMRC50174.2021.9569716).
- 唐 伦: 男, 教授, 博士生导师, 研究方向为新一代无线通信网络、异构蜂窝网络、软件定义无线网络等。
- 文明艳: 女, 硕士生, 研究方向为移动边缘计算辅助智能驾驶技术、联邦学习效率优化等。
- 单贞贞: 女, 硕士生, 研究方向为边缘智能协同计算资源分配、联邦学习资源协同优化等。
- 陈前斌: 男, 教授, 博士生导师, 研究方向为个人通信、多媒体信息处理与传输、下一代移动通信网络、异构蜂窝网络等。

责任编辑: 余 蓉