中国科学 : 信息科学 2025 年 第 55 卷 第 6 期: 1508–1521
SCIENTIA SINICA Informationis
论文

面向多用户语义知识库高效可信联邦遗忘方案

林怡静[1], 任金科[2,4*], 王嘉诚[5*], 马楠[1,3], 张晔[6], 崔曙光[4,2]

1. 北京邮电大学网络与交换技术全国重点实验室, 北京 100876, 中国
2. 深圳市未来智联网络研究院, 深圳 518172, 中国
3. 鹏城实验室宽带通信部, 深圳 518055, 中国
4. 香港中文大学 (深圳) 理工学院与广东省未来智联网络重点实验室, 深圳 518172, 中国
5. College of Computing and Data Science, Nanyang Technological University, Singapore 639798, Singapore
6. 北京信息科技大学计算机学院, 北京 100192, 中国

* 通信作者. E-mail: jinkeren@cuhk.edu.cn, jiacheng.wang@ntu.edu.sg

收稿日期: 2024–10–14; 修回日期: 2024–12–29; 接受日期: 2025–02–17; 网络出版日期: 2025–06–10

博士后创新人才支持计划 (批准号: BX20240045)、北京市自然科学基金 (批准号: QY24203)、河套深港科技合作区基础研究项目 (批准号: HZQB-KCZYZ-2021067)、国家自然科学基金 (批准号: 62293482)、深圳市杰出人才培养基金 (批准号: 202002)、广东省科研项目 (批准号: 2017ZT07X152, 2019CX01X104)、广东省未来智联网络重点实验室 (批准号: 2022B1212010001) 和深圳市大数据和人工智能重点实验室 (批准号: ZDSYS201707251409055) 资助项目

摘要 语义通信通过发送方和接收方共享语义知识库, 从原始数据中提取核心语义信息, 并传输编码后的语义信息, 从而减少传输数据量, 提高传输效率. 在多用户语义通信中, 语义知识库依赖多用户数据训练以实现精准的语义理解和传输. 为保护用户隐私, 联邦学习允许多用户在不共享原始数据的前提下协作训练语义知识库. 然而, 在协同训练过程中可能引入过时、不相关、有害等无效数据, 影响多用户语义知识库的可靠性和整体通信性能. 为消除无效数据的影响并维持语义知识库的可用性, 需要消耗大量通信和计算资源重新协同训练语义知识库. 为解决该问题, 本文创新地提出一种面向多用户语义知识库的高效可信联邦遗忘方案, 考虑单层多用户语义通信和多层多用户语义通信的不同层次需求, 设计基于梯度上升和可信执行环境的联邦遗忘方法, 在有效移除无效数据影响的同时确保语义知识库的可靠性. 实验结果表明, 在单层和多层的语义通信场景下, 提出方案大幅减少计算开销, 且在遗忘数据的识别准确率和遗忘后语义知识库的准确率上与重训练方法相当.

关键词 多用户语义通信, 语义知识库, 联邦遗忘

1 引言

根据国际数据公司 (International Data Cooperation) 的预测, 全球数据量将在 2025 年达到 175 泽字节 (ZB) [1]. 同时, 随着高清视频流媒体、自动驾驶、智能制造等复杂应用场景的扩展 [2], 现有通信系统在高效性和智能性方面表现出明显局限, 其性能已逼近经典信息论所提出的理论极限 [3]. 与传统通信需要传输完整的原始数据不同, 语义通信通过发送方和接收方共享语义知识库并传递核心语义信

引用格式: 林怡静, 任金科, 王嘉诚, 等. 面向多用户语义知识库高效可信联邦遗忘方案. 中国科学: 信息科学, 2025, 55: 1508–1521, doi: 10.1360/SSI-2024-0304
Lin Y J, Ren J K, Wang J C, et al. Efficient and trusted federated unlearning for multi-user semantic knowledge base. Sci Sin Inform, 2025, 55: 1508–1521, doi: 10.1360/SSI-2024-0304

© 2025《中国科学》杂志社

https://www.sciengine.com/doi/10.1360/SSI-2024-0304

息, 在去除冗余信息的同时减少传输数据量, 降低带宽需求和计算负担. 其中, 语义知识库是一种具备结构化特征的知识网络模型, 它能够从数据中抽取语义信息, 提供语义知识描述, 存储和记忆语义信息 [4].

语义知识库是体现语义通信智能性和高效性的重要组成部分 [3]. 以端到端语义通信为例, 发送方利用共享的语义知识库, 从原始数据中提取语义信息, 对其进行编码压缩, 仅传输核心语义内容; 接收方则通过解码重建语义信息, 确保接收到的信息准确反映发送方的意图, 实现高效的语义传输 [5]. 然而, 端到端的语义通信模式难以应对多用户间复杂的语义信息共享和协作需求. 多用户语义通信通过联邦学习共享和协同更新语义知识库 [6], 允许多个用户在保护隐私的同时进行高效的语义信息交换和协作, 进一步提升系统的智能性和通信效率. 具体包括: 在本地使用语义通信过程中可能传输或已传输的数据训练本地语义知识库, 并共享更新后的模型参数. 这些模型参数通过中央服务器进行聚合, 更新全局语义知识库, 从而在保护用户隐私的同时, 确保共享语义知识库的准确性与一致性 [7].

然而, 在多用户语义通信过程中, 用户在进行本地训练时可能引入过时、不相关、有害等无效数据, 影响语义知识库的性能和可靠性. 例如, 在基于多用户语义通信的虚拟现实交互过程中, 恶意用户可能在训练过程中引入虚假场景或错误的交互数据, 导致全局共享的语义知识库生成错误的虚拟场景信息, 影响其他用户体验, 甚至破坏整个虚拟环境的运行稳定性. 因此, 如何有效地消除无效数据的影响, 同时维持语义知识库的可用性, 成为当前亟待解决的问题. 现有研究采用直接重新训练语义知识库的方式. 这种方法虽然可以恢复语义知识库的准确性, 但计算开销极高, 且难以支持高效更新.

针对上述问题, 本文提出面向多用户语义知识库的高效可信遗忘方案, 综合考虑语义知识库的可用性和消除无效数据影响, 涵盖单层多用户语义通信和多层跨域语义通信两种场景, 在无需重新训练的前提下从全局知识库中移除无效数据影响, 可应用于低空经济等场景, 提升多用户语义通信的可靠性. 本文的主要贡献包括以下三点:

(1) 针对无效数据对语义知识库的影响, 提出基于联邦遗忘的多用户语义知识库更新框架, 在无需重训练的前提下, 有效消除过时、不相关、有害等无效数据的影响, 保证语义知识库的准确性和可靠性;

(2) 面向单层多用户语义通信场景, 提出基于补偿梯度上升和可信执行环境的单层语义知识库遗忘方法, 在利用梯度上升遗忘无效数据后执行补偿训练, 结合可信执行环境确保无效数据移除后的语义知识库的准确性和效率;

(3) 面向多层跨域语义通信场景, 针对无效数据影响在不同层次间传播导致语义知识库受影响范围增大, 提出基于梯度上升和历史梯度的多层语义知识库遗忘方法, 通过梯度上升和回溯历史梯度信息, 消除各层级语义知识库中的无效数据影响.

本文其余章节结构如下: 第 2 节从语义知识库和联邦遗忘两个角度阐述相关工作; 第 3 节构建基于联邦遗忘的多用户语义知识库更新框架和工作流程; 第 4 节给出提出方案的理论分析和实验结果, 验证提出方案的计算开销、可用性和遗忘有效性; 第 5 节总结全文.

2　相关工作

2.1　语义通信与语义知识库

语义知识库是语义通信区别于传统通信的关键模块, 具备语义理解、知识表示和语义推理功能, 可以为语义通信系统提供全局共享的基础知识, 确保语义传输的准确性和高效性. 当前, 针对语义知识库的研究主要关注语义知识构建与优化和多用户语义通信两个方面.

在语义知识库构建与优化方面, 文献 [4] 通过多层级语义知识库设计、语义知识库动态演进和多智能体知识库协同更新构建高效的语义知识库. 文献 [5] 提出面向语义通信的知识库体系化设计, 通

https://www.sciengine.com/doi/10.1360/SSI-2024-0304

过语义知识库融合信源 – 信道联合编解码、跨协议语义通信、边缘缓存与部署、通信 – 计算资源联合管理等方式提升语义通信的智能性和高效性. 文献 [8] 提出在有限知识库条件下, 基于归一化条件复杂度的语义压缩条件估计方法, 估计在有无知识库的情况下特定数据集消息编码所需的额外复杂度. 文献 [9] 提出语义知识库驱动的生成式语义通信架构, 通过信源知识库、任务知识库和信道知识库三个子语义知识库实现高效的语义编码和语义传输.

在多用户语义通信方面, 文献 [7] 提出基于联邦学习的多用户语义通信系统部署方法, 通过用户在本地数据上训练共享知识库, 在无需共享原始数据的前提下实现模型参数的共享, 构建全局语义知识库, 在保护用户隐私的同时提升多用户语义通信的效率. 文献 [10] 面向多用户语义通信提出基于客户端性能聚合的联邦学习机制, 通过减少每轮训练中传输的信息量来节省带宽, 在不影响用户隐私的前提下充分利用客户端数据, 提升语义通信效率. 文献 [11] 提出基于联邦知识蒸馏的语义知识库同步框架, 通过蒸馏机制学习异构全局知识库, 同时压缩全局知识库以提高同步效率, 在提升准确性的同时减少通信开销.

然而, 上述研究未能解决过时、不相关、有害数据对训练语义知识库带来的负面影响问题, 导致语义知识库长期积累无效数据, 影响整个语义通信系统的可靠性. 因此, 如何高效、可信地消除无效数据的影响, 成为提升语义知识库可靠性的关键.

2.2　基于事前容错的联邦学习

在多用户语义通信中, 不同用户通过联邦学习共享和协同训练语义知识库, 以便在通信过程中高效提取和传递语义信息. 具体来说, 它通过在客户端上训练语义知识库模型, 并共享模型参数或梯度, 而非原始数据, 实现数据 “可用不可见”. 由中心聚合节点整合本地模型更新, 从而生成全局模型. 然而, 客户端可能上传无效模型参数, 利用这些数据训练全局语义知识库会影响系统的整体性能. 为此, 现有方法主要采用事前容错机制, 包括客户端质量筛选和模型鲁棒聚合, 以减少无效参数的负面影响.

在客户端质量筛选方面, 文献 [12] 通过客户端独立训练的相关数据选择器为本地数据动态分配相关性评分, 并根据评分筛选对全局模型有贡献的数据子集, 服务器使用验证集评估并反馈客户端的贡献. 文献 [13] 逐步增加客户端数量提升数据多样性, 同时结合当前和历史轮次的性能指标评估客户端贡献, 利用权重衰减动态调整评估重点. 文献 [14] 收集客户端的状态信息, 利用强化学习训练的策略网络动态选择参与任务的客户端, 根据全局模型表现生成奖励信号, 优化客户端选择策略.

在模型鲁棒聚合方面, 文献 [15] 提出拜占庭 (Byzantine) 容错的鲁棒梯度下降方法, 通过对参与者的本地模型权重进行排序和选择以过滤掉异常模型权重. 该方法通过计算本地模型权重向量之间的欧氏距离, 结合每个参与者与其他固定数量参与者模型权重的距离和, 选取距离总和最小的模型作为聚合结果. 文献 [16] 提出高效的模型聚合方法, 通过迭代剔除上传梯度中的离群值来抵抗拜占庭攻击. 首先计算所有上传梯度的均值, 找到与均值差距最大的梯度并将其剔除. 重复这一过程, 直到剔除的梯度数量达到假定的拜占庭节点比例. 然后对剩余的梯度重新计算均值作为最终聚合梯度, 并根据该聚合梯度更新模型权重. 文献 [17] 提出基于采样的模型聚合方法, 通过计算模型采样向量之间的欧几里得 (Euclidean) 距离, 与其他模型距离较大的模型很可能来自恶意客户端.

2.3　基于事后容错的联邦遗忘

联邦遗忘是解决联邦学习事后容错的补充计算范式. 基于事前容错的联邦学习方法难以完全筛选和移除无效数据, 导致带有无效数据影响的模型参数影响全局模型. 联邦遗忘通过允许全局聚合模型在不重新训练整个模型的情况下, 精确地遗忘这些数据, 从而消除它们对模型的负面影响, 提升模型的可靠性. 当前, 针对联邦遗忘的研究主要关注联邦遗忘方法设计与优化和隐私保护与激励机制两个方面.

在联邦遗忘方法设计与优化方面, 文献 [18] 首次提出基于历史参数更新的联邦遗忘框架, 通过校

https://www.sciengine.com/doi/10.1360/SSI-2024-0304

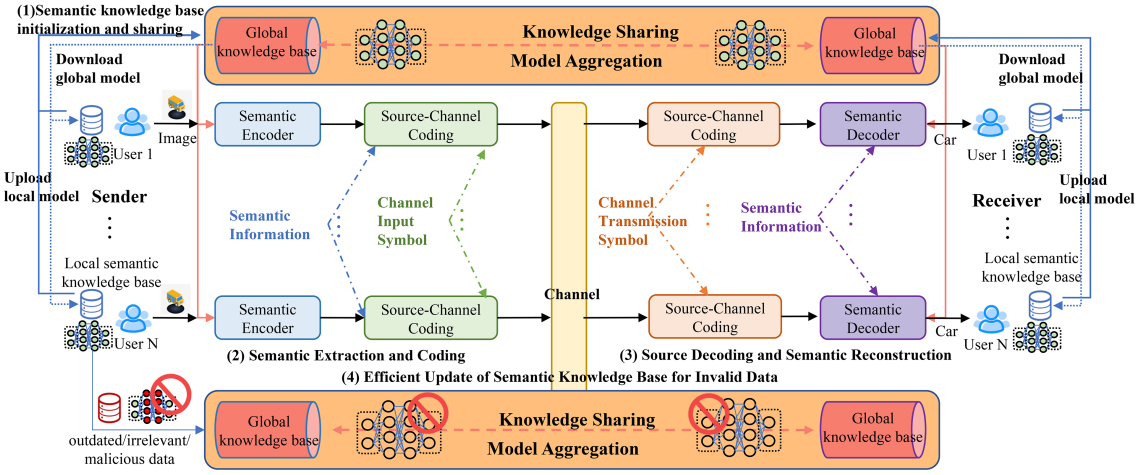


图 1　(网络版彩图) 多用户语义知识库.

Figure 1　(Color online) Multi-user semantic knowledge base.

准保留的更新来加速模型重构, 能够快速从全局模型中移除特定客户端的数据影响, 显著减少构建遗忘模型所需的时间. 文献 [19] 对联邦遗忘进行形式化定义, 并提出快速重训练的方法, 能够完全从训练好的联邦学习模型中删除数据样本, 允许数据持有者在不共享本地数据的情况下高效地进行遗忘操作. 文献 [20] 提出基于隔离分片和编码计算的可扩展联邦遗忘框架, 通过减少受影响的客户端数量和压缩模型参数来优化存储与计算效率. 文献 [21] 提出基于聚类聚合的异步联邦遗忘机制, 通过限制重新训练的范围, 显著减少了重新训练的计算开销, 并优化了客户端簇的分配策略.

在隐私保护与激励机制方面, 文献 [22] 提出基于编码计算的去中心化联邦遗忘框架, 保障遗忘过程的可扩展性与安全性. 文献 [23] 结合四阶段博弈模型设计联邦学习和遗忘的激励机制, 通过将用户的多维私有信息简化为一维指标以优化激励设计, 鼓励有价值的用户继续参与遗忘过程. 文献 [24] 提出基于双层激励的联邦遗忘机制, 首先利用进化博弈论模型化动态参与过程, 吸引剩余客户端参与重新训练任务, 然后结合深度强化学习, 动态选择剩余客户端参与遗忘过程, 以减少数据不平衡带来的偏差.

然而, 上述研究未能解决多层联邦遗忘的问题, 因为在多层结构中, 无效数据的影响会被放大. 同时, 尚未有研究考虑将联邦遗忘应用于多用户语义通信, 以解决语义知识库中无效数据积累的问题. 因此, 研究面向多用户语义知识库的高效可信遗忘方案具有重要的研究意义和实际应用价值.

3　面向多用户语义知识库的高效可信遗忘方案

3.1　基于联邦遗忘的多用户语义知识库更新框架

在基于深度学习的语义通信中, 语义知识库是经过训练/优化的参数模型, 通过发送方和接收方事先共享这些参数, 以确定语义编解码器中的具体参数配置 [4,7]. 如图 1 所示, 本文考虑多用户语义通信, 每个用户均可能成为信息发送方或接收方, 表示为集合 $\mathcal{N}=\{1,\ldots,N\}$, 总计 N 个用户, 通过共享语义知识库、语义提取和信源编码、信源解码和语义重构以及语义知识库的高效更新, 以提高系统的通信效率和准确性.

(1) 语义知识库的初始化和共享: 每个用户首先通过联邦学习协同训练和更新全局语义知识库, 定义语义知识库全局参数为 $\boldsymbol{w}$, 全局损失函数为 $F(\boldsymbol{w})$. 在每一轮训练中, 选定一个用户作为聚合节点以协调各用户. 在第 k 轮训练时, 聚合节点将当前的全局知识库参数 $\boldsymbol{w}^k$ 发送给所有用户. 每个用户 n 根据其本地数据集 $\mathcal{D}_n$ 进行本地训练, 最小化其本地损失函数 $F_n(\boldsymbol{w}^k)$, 得到本地更新后的参数 $\boldsymbol{w}_n^k$.

https://www.sciengine.com/doi/10.1360/SSI-2024-0304

随后, 所有用户将本地参数 $\boldsymbol{w}_n^k$ 反馈给聚合节点, 聚合节点通过联邦平均算法聚合本地参数, 得到下一轮全局参数 $\boldsymbol{w}^{k+1}$, 即 $\boldsymbol{w}^{k+1}=\frac{1}{N}\sum_{n=1}^{N}\boldsymbol{w}_n^k$. 接着, 聚合节点将新的全局参数 $\boldsymbol{w}^{k+1}$ 共享给所有用户, 开启下一轮协同训练. 通过 K 轮迭代更新, 语义知识库得以不断优化, 从而提升多用户语义通信系统的整体性能.

(2) 语义提取和信源编码: 发送方基于共享的语义知识库 $\boldsymbol{w}$, 使用语义编码器 $\boldsymbol{T}_{\theta_e}$ (其中 θ_e 为语义编码器的参数), 从原始数据 (例如图片, 记作 $\boldsymbol{r}$) 中提取语义信息 $\boldsymbol{x}$, 该过程表示为 $\boldsymbol{x}=\boldsymbol{T}_{\theta_e}(\boldsymbol{r},\boldsymbol{w})$. 随后, 联合信源信道编码器将语义信息映射为信道输入符号, 并通过无线信道进行传输, 表示为 $\boldsymbol{y}=\boldsymbol{c}\cdot\boldsymbol{x}+\sigma$, 其中 $\boldsymbol{c}$ 表示信道增益系数, σ 为服从正态分布的高斯 (Gauss) 噪声.

(3) 信源解码和语义重构: 接收方接收到信道传输的信号 $\boldsymbol{y}$, 通过联合信源信道解码器输出语义信息 $\hat{x}$. 接着, 接收方利用共享的语义知识库 $\boldsymbol{w}$, 使用语义解码器 $\boldsymbol{R}_{\theta_d}$ (其中 θ_d 为语义解码器的参数) 对语义信息 $\hat{x}$ 进行重构, 恢复出发送方的意图 $\hat{r}$, 该过程表示为 $\hat{r}=\boldsymbol{R}_{\theta_d}(\hat{x},\boldsymbol{w})$.

(4) 面向无效数据的语义知识库的高效更新: 随着通信环境、数据源和用户需求的变化, 多用户语义知识库中的部分用户数据可能变得过时、不相关, 甚至包含有害数据. 由于语义知识库共享和初始化过程无法访问用户的原始数据, 且语义知识库训练过程中存在参数更新的随机性, 用户数据与模型参数的更新难以直接对应 [25], 使得直接定位并消除语义知识库中无效数据的影响变得非常困难. 此外, 随着多轮训练的累积, 无效数据的负面影响可能会逐渐放大, 进一步降低模型的性能. 为解决上述问题, 面向单层多用户语义通信场景, 本文提出基于补偿梯度上升和可信执行环境的单层语义知识库遗忘方法, 通过有效遗忘无效数据对模型参数的影响和利用可信执行环境确保遗忘过程的安全性, 在不重新训练的情况下移除无效数据的影响, 如第 3.2 小节所述; 面向多层跨域语义通信场景, 本文提出基于梯度上升和历史梯度的多层语义知识库遗忘方法, 结合历史梯度信息和动态调整模型参数以修正无效数据的影响, 有效解决多层结构中无效数据影响放大的问题, 如第 3.3 小节所述.

3.2　基于补偿梯度上升和可信执行环境的单层语义知识库遗忘方法

为有效应对无效数据对语义知识库的负面影响, 特别是在多用户共享语义知识库中需要遗忘特定用户的数据时, 本文提出基于补偿梯度上升和可信执行环境的遗忘机制. 假设需要遗忘的用户集合为 $\mathcal{N}_u\subset\mathcal{N}$, 其对应的本地数据为 $\mathcal{D}_{N_u}$. 理想情况下, 遗忘前后语义知识库应满足

$$\begin{cases} I\left(\boldsymbol{w}(\mathcal{D}_{N_u});\boldsymbol{w}_u(\mathcal{D}_{N_u})\right)=0, \\ I\left(\boldsymbol{w}(\mathcal{D}-\mathcal{D}_{N_u});\boldsymbol{w}_u(\mathcal{D}-\mathcal{D}_{N_u})\right)=1, \end{cases} \tag{1}$$

其中, I 表示遗忘前后语义知识库在遗忘数据 $\mathcal{D}_{N_u}$ 和剩余数据 $\mathcal{D}-\mathcal{D}_{N_u}$ 上的互信息. 第一个条件表明遗忘后的语义知识库 $\boldsymbol{w}_u$ 不包含遗忘数据的信息, 即其互信息应为 0; 第二个条件确保遗忘过程不会影响剩余数据的知识库更新, 即两者在剩余数据集上的互信息应保持为 1. 上述两个条件的目标是在进行有效遗忘操作的同时维持语义知识库的可用性.

现有方法多数由聚合节点发起遗忘请求要求所有剩余用户参与遗忘过程, 导致计算和通信开销大. 本文探讨由遗忘用户主动发起遗忘过程. 首先, 需要从历史更新中移除遗忘用户 $\mathcal{N}_u$ 的历史更新并聚合成新的全局模型作为初始遗忘模型, 表示为 $\boldsymbol{w}_u^0=\frac{1}{N-N_u}\sum_{n\in\mathcal{N}-\mathcal{N}_u}\boldsymbol{w}_n^{k+1}$. 当遗忘用户发现之前用于训练语义知识库的无效数据时, 可以在初始遗忘模型的基础上通过本地执行梯度上升操作来移除无效数据对语义知识库的影响. 梯度上升的目标是增加损失函数, 从而减小无效数据对语义知识库的贡献, 表示为

$$\boldsymbol{w}_u^{k+1}=\boldsymbol{w}_u^k+\eta_u\nabla_{\boldsymbol{w}}F(\boldsymbol{w}_u^k,\mathcal{D}_{N_u}), \tag{2}$$

其中 $\boldsymbol{w}_u^k$ 是第 k 轮中遗忘用户的本地语义知识库参数, η_u 是学习率, $\nabla_{\boldsymbol{w}}F(\boldsymbol{w}_u^k,\mathcal{D}_{N_u})$ 表示遗忘数据 $\mathcal{D}_{N_u}$ 的损失函数梯度. 由于遗忘用户在全局语义知识库 $\boldsymbol{w}$ 的基础上执行梯度上升操作, 虽然能够有效

https://www.sciengine.com/doi/10.1360/SSI-2024-0304

遗忘无效数据的影响, 但同时也会导致语义知识库的可用性下降, 即对其他未遗忘用户数据的准确性降低. 为解决该问题, 遗忘后的语义知识库 $\boldsymbol{w}_u^{k+1}$ 需要进行补偿训练, 以恢复语义知识库的可用性. 具体而言, 补偿过程可以通过选定部分用户进行轻量化的本地训练, 利用这些用户的数据来重新调整语义知识库. 假设补偿用户集合为 $\mathcal{N}_c \subseteq \mathcal{N}-\mathcal{N}_u$, 总计 N_c 个补偿用户, 其本地数据集为 $\mathcal{D}_{N_c} \subseteq \mathcal{D}-\mathcal{D}_{N_u}$, 通过补偿用户执行本地梯度下降操作来最小化全局损失函数, 表示为

$$\boldsymbol{w}_c^{k+1} = \boldsymbol{w}_c^k - \eta_c \nabla_{\boldsymbol{w}} F(\boldsymbol{w}_c^k, \mathcal{D}_{N_c}), \tag{3}$$

其中 $\boldsymbol{w}_c^k$ 是第 k 轮补偿用户的本地语义知识库, η_c 是补偿训练的学习率, $\nabla_{\boldsymbol{w}} F(\boldsymbol{w}_c^k, \mathcal{D}_{N_c})$ 表示补偿用户数据集 $\mathcal{D}_{N_c}$ 损失函数梯度. 随后, 补偿用户聚合作为遗忘后的最终语义知识库 $\boldsymbol{w}^{k+1} = \frac{1}{N_c}\sum_{c=1}^{N_c} \boldsymbol{w}_c^k$. 在不需要所有剩余用户参与的情况下, 这种补偿机制有效降低了系统开销, 同时保持了语义知识库可用性.

由于遗忘用户执行梯度上升操作可能导致语义知识库的可用性下降, 因此难以判别该过程是正常的遗忘操作, 还是用户故意上传恶意参数以破坏语义知识库. 因此, 本文面向遗忘用户提出基于可信执行环境的可信遗忘方法, 隔离和保护遗忘过程中的计算, 防止用户篡改参数或进行恶意操作, 同时在保证模型遗忘无效数据的前提下维持语义知识库的完整性和安全性. 可信执行环境由硬件厂商或系统架构管理者提供, 参与用户依赖该环境进行安全计算和验证操作, 包括密钥生成、签名、共享、验证和认证等关键过程. 具体如下.

遗忘用户在可信执行环境中生成一对公私钥, 用于签名和验证. 公钥为 PK_u, 私钥为 SK_u, 密钥生成过程表示为 $(\mathrm{PK}_u, \mathrm{SK}_u) = \mathsf{KeyGen}()$. 遗忘用户在可信环境中执行梯度上升操作后, 使用私钥对结果签名, 用于证明遗忘操作合法性. 由于签名操作在可信执行环境内部, 不会暴露私钥, 表示为 $\sigma_u = \mathsf{Sign}(\mathrm{SK}_u, \boldsymbol{w}_u^{k+1})$. 遗忘用户通过安全通道 (由可信执行环境提供者的硬件或软件机制支持) 将公钥和签名安全地共享给剩余用户. 剩余用户在接收到更新的模型参数和签名后, 利用公钥验证签名的真实性, 验证过程表示为 $\mathsf{Verify}(\mathrm{PK}_u, \boldsymbol{w}_u^{k+1}, \sigma_u) = 1/0$. 为验证遗忘操作是否在可信执行环境中执行, 通过可信执行环境提供者提供的 API 和硬件支持来进行远程认证生成可信的认证信息, 以验证执行正确性, 表示为 $\sigma_u = \mathsf{Attest}\left(\mathcal{H}(\boldsymbol{w}_u^{k+1})\right)$, 其中 Attest 表示远程认证过程, $\mathcal{H}$ 表示遗忘后的模型参数 $\boldsymbol{w}_u^{k+1}$ 的哈希 (Hash) 值.

3.3　基于梯度上升和历史梯度的多层语义知识库遗忘方法

在多层跨域语义通信中, 语义知识库的初始化和共享依赖多个层级用户的数据协同训练. 以智慧城市为例, 多层次的语义知识库框架可分为用户层、边缘层、区域层和中央聚合节点. 用户层对应智能设备, 边缘层为社区网络的基站, 区域层汇总城市内的所有社区网络基站, 中央聚合节点则负责省级范围内的语义信息汇总, 实现跨地域、跨层次的大规模数据处理与语义传输. 该架构可以看作是一个分层的聚合框架, 每一层的语义知识库更新都依赖下层的汇总结果. 用户层负责生成原始语义信息, 边缘层执行初步聚合, 区域层整合边缘节点的更新, 最终在中央节点完成全局语义知识库的更新.

在初始化阶段, 假设有 N 个用户, 每个用户 n 基于本地数据集 $\mathcal{D}_n$ 更新本地知识库, 表示为 $\boldsymbol{w}_n^{k+1} = \boldsymbol{w}_{\mathrm{central}}^k - \eta \nabla_{\boldsymbol{w}} F(\boldsymbol{w}_{\mathrm{central}}^k, \mathcal{D}_n)$, 其中 $\boldsymbol{w}_{\mathrm{central}}^k$ 为上轮从中央节点同步的全局语义知识库参数, $\nabla_{\boldsymbol{w}} F(\boldsymbol{w}_{\mathrm{central}}^k, \mathcal{D}_n)$ 为局部梯度. 边缘层有 J 个基站, 每个基站 j 汇总其管辖用户的本地更新, 初步聚合语义信息, 表示为 $\boldsymbol{w}_{\mathrm{edge},j}^{k+1} = \frac{1}{N}\sum_{n\in\mathcal{N}_{\mathrm{edge},j}} \boldsymbol{w}_{n,\mathrm{local}}^{k+1}$, 其中 $\mathcal{N}_{\mathrm{edge},j}$ 是边缘节点 j 负责的用户集合. 区域层进一步汇总不同边缘节点的聚合结果, 最终在中央节点进行全局聚合. 区域层和中央节点的聚合过程类似边缘层, 分别得到区域层的聚合结果和中央节点的最终全局更新 $\boldsymbol{w}_{\mathrm{region},r}^{k+1}$ 和 $\boldsymbol{w}_{\mathrm{central},c}^{k+1}$, $\mathcal{N}_{\mathrm{region},r}$ 是区域层 r 下属的边缘节点集合, $\mathcal{N}_{\mathrm{central}}$ 是中央节点管理的所有区域层, 经过 K 轮迭代更新, 得到全局模型 $\boldsymbol{w}_{\mathrm{central}}^K$. 尽管多层次架构提升了处理大规模数据的效率和准确性, 但是随着用户范围扩大, 过时、不相关或有害数据的影响也会逐步放大. 为此, 在第 3.2 小节的基础上, 本文提出基于梯度上升

https://www.sciengine.com/doi/10.1360/SSI-2024-0304

和历史梯度的遗忘机制, 融合梯度上升与各层的历史信息, 有效削弱无效数据对全局模型的负面影响.

首先, 需要从用户集合中移除遗忘用户的历史更新, 然后经由层层聚合形成初始遗忘模型, 表示为

$$\boldsymbol{w}_u^0 = \frac{1}{|\mathcal{N}_{\text{central}}|}\sum_{r\in\mathcal{N}_{\text{central}}}\frac{1}{|\mathcal{N}_{\text{region},r}|}\sum_{j\in\mathcal{N}_{\text{region},r}}\frac{1}{|\mathcal{N}_{\text{edge},j}-\mathcal{N}_u|}\sum_{n\in\mathcal{N}_{\text{edge},j}-\mathcal{N}_u}\boldsymbol{w}_{n,\text{local}}^{k+1}. \tag{4}$$

随后, 遗忘用户基于式 (2), 在初始遗忘模型 $\boldsymbol{w}_u^0$ 的基础上执行梯度上升操作, 生成遗忘后的模型 $\boldsymbol{w}_u^k$, 以减弱无效数据的影响. 如第 3.2 小节所述, 为防止遗忘用户或者其他层节点故意上传恶意参数破坏语义知识库, 采用可信执行环境进行安全计算和验证操作, 维持语义知识库的安全性. 由于多层次结构中无效数据的影响更为显著 (见第 4.3 小节实验结果), 与单层不同, 补偿用户 $\mathcal{N}_c$ 首先在最新遗忘模型上执行梯度下降来最小化全局函数, 其他用户依然使用未更新的遗忘模型 $\boldsymbol{w}_u^k$, 层层聚合后的全局模型可以表示为

$$\boldsymbol{w}_c^{k+1} = \sum_{r\in\mathcal{N}_{\text{central}}}\sum_{j\in\mathcal{N}_{\text{region},r}}\frac{\hat{\boldsymbol{w}}}{|\mathcal{N}_{\text{central}}||\mathcal{N}_{\text{region},r}|}, \tag{5}$$

其中, $\hat{\boldsymbol{w}}$ 表示 $(\sum_{n\in\mathcal{N}_{\text{edge},j}-\mathcal{N}_c-\mathcal{N}_u}\boldsymbol{w}_u^k+\sum_{c\in\mathcal{N}_c}(\boldsymbol{w}_u^k-\eta_c\nabla_{\boldsymbol{w}}F(\boldsymbol{w}_u^k,\mathcal{D}_{N_c})))/|\mathcal{N}_{\text{edge},j}-\mathcal{N}_u|$. 然后, 在中央聚合节点处利用历史更新校准遗忘模型, 表示为

$$\boldsymbol{w}_u^{k+1} = \boldsymbol{w}_c^{k+1} + \left\|\frac{1}{N-N_u}\sum_{n=1}^{N-N_u}(\boldsymbol{w}_n^{k+1}-\boldsymbol{w}_{\text{central}}^K)\right\|\frac{\hat{\boldsymbol{w}}-\boldsymbol{w}_c^{k+1}}{\|\hat{\boldsymbol{w}}-\boldsymbol{w}_c^{k+1}\|}. \tag{6}$$

3.4　收敛性分析

在非独立同分布数据上进行单层语义知识库遗忘方法的收敛性分析, 以获得遗忘后全局模型与最优模型之间差距. 为此, 本文提出了以下假设 [26].

假设1　每个剩余用户的目标函数对任意两个模型 $\boldsymbol{w}$ 和 $\boldsymbol{v}$ 是 μ-strong 强凸的, 即

$$F_n(\boldsymbol{w})-F_n(\boldsymbol{v}) \leqslant (\boldsymbol{w}-\boldsymbol{v})^{\mathrm{T}}\nabla F_n(\boldsymbol{v})+\frac{\mu}{2}||\boldsymbol{w}-\boldsymbol{v}||^2. \tag{7}$$

假设2　每个剩余用户的目标函数对任意两个模型 $\boldsymbol{w}$ 和 $\boldsymbol{v}$ 是 L-smooth 平滑的, 即

$$||\nabla F_n(\boldsymbol{v})-\nabla F_n(\boldsymbol{w})|| \leqslant L||\boldsymbol{v}-\boldsymbol{w}||. \tag{8}$$

基于上述假设和文献 [26], 可以推导出下一轮模型 $\boldsymbol{w}_c^{k+1}$ 与全局最优模型 $\boldsymbol{w}^*$ 之间的差距,

$$\mathbb{E}||\boldsymbol{w}_c^{k+1}-\boldsymbol{w}^*||^2 = (1-\eta_c\mu)\mathbb{E}||\boldsymbol{w}_c^{k+1}-\boldsymbol{w}^*||^2+\eta_c^2 B, \tag{9}$$

其中 B 是常数. 令 $\widetilde{\boldsymbol{w}}^*$ 表示最优的全局遗忘模型, 根据文献 [23], 选择学习率 $\eta_c=\frac{\epsilon}{k+1}$, 其中 $\epsilon\geqslant\frac{1}{2\mu}$, 可以得到期望差距上界,

$$\mathbb{E}||\boldsymbol{w}_c^{k+1}-\widetilde{\boldsymbol{w}}^*||^2 \leqslant (1-\eta_c\mu)||\boldsymbol{w}^*-\widetilde{\boldsymbol{w}}^*||^2+B. \tag{10}$$

根据 $\boldsymbol{w}^*$ 和 $\widetilde{\boldsymbol{w}}^*$ 的定义可以得到

$$\sum_{n\in\mathcal{N}\setminus\mathcal{N}_u}(\nabla F_n(\boldsymbol{w}^*)-\nabla F_n(\widetilde{\boldsymbol{w}}^*)) = \sum_{n\in\mathcal{N}}\nabla F_n(\boldsymbol{w}^*)-\sum_{n\in\mathcal{N}\setminus\mathcal{N}_u}(\nabla F_n(\boldsymbol{w}^*)+\nabla F_n(\widetilde{\boldsymbol{w}}^*)) = -\sum_{n\in\mathcal{N}}\nabla F_n(\boldsymbol{w}^*). \tag{11}$$

随后, 推导 $||\boldsymbol{w}^*-\widetilde{\boldsymbol{w}}^*||$ 下界, 表示为

$$\left\|\sum_{n\in\mathcal{N}}\nabla F_n(\boldsymbol{w}^*)\right\| \overset{(11)}{=} \sum_{n\in\mathcal{N}\setminus\mathcal{N}_u}(\nabla F_n(\boldsymbol{w}^*)-\nabla F_n(\widetilde{\boldsymbol{w}}^*)) \overset{(8)}{\leqslant} \sum_{n\in\mathcal{N}\setminus\mathcal{N}_u}L||\boldsymbol{w}^*-\widetilde{\boldsymbol{w}}^*||. \tag{12}$$

https://www.sciengine.com/doi/10.1360/SSI-2024-0304

然后, 推导 $||\boldsymbol{w}^* - \widetilde{\boldsymbol{w}}^*||$ 上界, 表示为

$$\begin{aligned} -\sum_{n\in\mathcal{N}} \nabla F_n(\boldsymbol{w}^*)^{\mathrm{T}}(\boldsymbol{w}^* - \widetilde{\boldsymbol{w}}^*) &\overset{(11)}{=} -\sum_{n\in\mathcal{N}} (\nabla F_n(\boldsymbol{w}^*) - \nabla F_n(\widetilde{\boldsymbol{w}}^*))^{\mathrm{T}}(\boldsymbol{w}^* - \widetilde{\boldsymbol{w}}^*) \\ &\overset{(7)}{\geqslant} \sum_{n\in\mathcal{N}\setminus\mathcal{N}_u} (F_n(\boldsymbol{w}^*) - F_n(\widetilde{\boldsymbol{w}}^*)) + \frac{\mu}{2}||\boldsymbol{w}^* - \widetilde{\boldsymbol{w}}^*|| \geqslant \frac{\mu}{2}||\boldsymbol{w}^* - \widetilde{\boldsymbol{w}}^*||. \end{aligned} \tag{13}$$

结合式 (10), (12) 和 (13), 可得

$$\mathbb{E}||\boldsymbol{w}^* - \widetilde{\boldsymbol{w}}^*||^2 \overset{(\mathrm{d})}{\leqslant} \frac{2(1-\eta_c\mu)(N_u-1)}{\mu} \sum_{n\in\mathcal{N}_u} ||\nabla F_n(\boldsymbol{w}^*)||^2 + B, \tag{14}$$

其中条件 (d) 表示遗忘过程中只剩余一个用户. 由于 η_c 随轮数 k 增大而减少, $||\boldsymbol{w}^* - \widetilde{\boldsymbol{w}}^*||$ 会逐渐减小, 因此模型将收敛.

4　实验评估

4.1　实验设置

本文在 MNIST [27], Fashion-MNIST [28] 和 CIFAR-10 [29] 3 个数据集上进行实验. 实验分别在独立同分布 (IID) 和非独立同分布 (Non-IID) 条件下进行. 在 Non-IID 条件中, 每个客户端有 80% 的数据样本属于一个主要类别, 其余 20% 的数据样本属于其他类别 [30]. MNIST 和 Fashion-MNIST 数据集使用相同的网络结构作为语义知识库: 两个 5×5 的卷积层, 结合 ReLU 激活函数和最大池化层来提取特征, 随后通过两层全连接层进行分类输出; CIFAR-10 数据集使用的网络包括两层卷积层和最大池化层, 随后通过三层全连接层实现分类任务.

本文提出方案 (Ours) 与以下 4 种方案进行对比: (1) FedRetrain (FR) [18] 通过在移除目标用户数据后对整个模型重新训练来实现遗忘, 虽然可以确保精确的遗忘效果, 但重新训练整个模型的计算和时间成本较高; (2) FedEraser (FE) [18] 采用 “以空间换时间” 的策略, 通过存储历史模型更新来实现高效的遗忘, 通过回溯存储的更新历史, 调整模型的参数; (3) RapidRetrain (RR) [19] 通过使用海森 (Hessian) 矩阵来加速遗忘过程, 海森矩阵可以帮助计算模型参数对目标数据的敏感度, 并有效移除该数据影响; (4) Shard [20] 通过将数据分片并隔离互不交互的用户, 在不显著降低模型性能的情况下, 缩小了遗忘操作的影响范围.

实验采用 3 个关键评估指标来全面衡量所提方案的有效性和性能表现: (1) 遗忘后模型的准确率, 即在执行遗忘操作后, 模型是否仍能有效完成任务, 直接反映方案的可用性; (2) 遗忘操作的时间开销, 指在尽可能短的时间内完成数据移除和模型调整, 以减少计算资源的消耗, 提高方案的可扩展性; (3) 基于成员推理攻击 [31] 的 F1 score, 通过观察模型对特定输入的预测结果, 推断该输入是否曾参与训练, 从而评估模型是否成功遗忘特定用户数据. 需要注意的是, F1 score 并非越低越好. 如果模型的准确率大幅下降, F1 score 自然也会降低, 这意味着模型整体性能退化. 因此, 3 个指标需要综合考虑. 如果某个方案能够在保持较高准确率且 F1 score 与重新训练相当的前提下, 显著缩短遗忘所需时间, 则该方案相比其他对比方案更具优势.

在语义知识库的初始化与共享阶段, 实验采用了 FedAvg [32] 联邦学习作为基础框架, 用户总数为 100, 学习率为 0.005, 批大小为 64, 本地训练轮次为 10, 全局训练轮次为 20 [18,20]. 在单层语义知识库遗忘中, 系统共有 100 个用户, 每轮训练中随机选取 20 个用户的数据进行协同训练. 在多层语义结构中, 随机选取 18 个用户的数据进行训练, 每 3 个用户组成一个边缘节点, 每 3 个边缘节点组成一个区域节点, 加上一个中央节点, 形成四层架构. 由于该模型缺乏有效的遗忘机制, 无法消除无效数据对模型的影响, 为后续提出的遗忘方案提供了对比基准.

https://www.sciengine.com/doi/10.1360/SSI-2024-0304

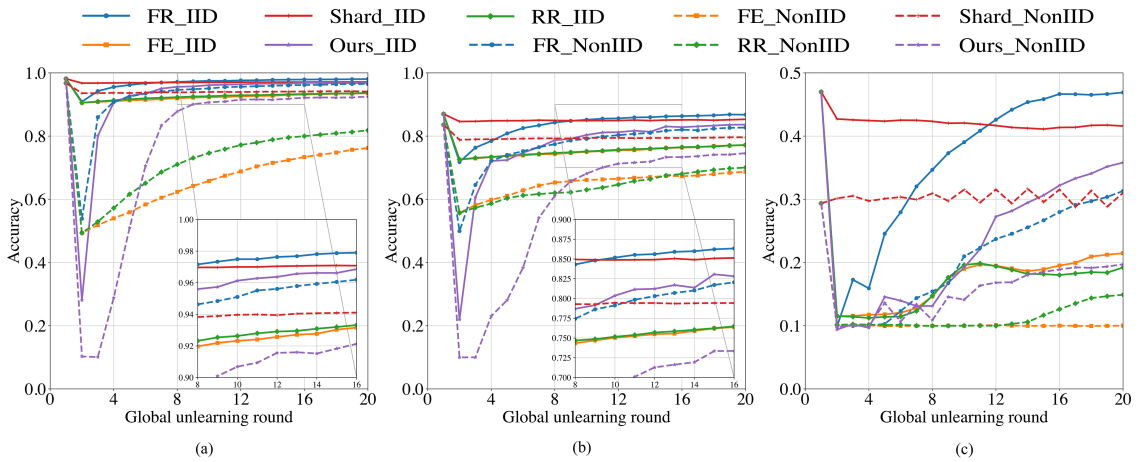


图 2　(网络版彩图) 单层语义知识库遗忘模型准确率. (a) MNIST; (b) Fashion-MNIST; (c) CIFAR-10

Figure 2　(Color online) Accuracy of the unlearned model for the single-layer semantic knowledge base. (a) MNIST; (b) Fashion-MNIST; (c) CIFAR-10.

4.2　单层语义知识库遗忘实验结果

在单层语义知识库遗忘任务中, 所有方法的全局遗忘轮次与全局训练轮次保持一致, 均为 20 次. 为降低计算开销, 各方法在执行遗忘操作时将本地校准轮次减少至 5 次. 实验中需要遗忘 600 个数据样本, 剩余用户需要进行校准, 以消除该遗忘用户的数据影响. 针对 Shard 方法, 将用户划分为 4 个分片, 每个分片包含 5 个用户. 对于提出的方法 (Ours), 梯度上升和补偿训练的总轮次为 20 次, 其中梯度上升执行 1 轮, 补偿训练执行 19 轮. 这种设置的原因在于, 梯度上升的目的是突出被遗忘数据对模型的影响, 而补偿训练则是为了平衡和消除这种影响. 因此, 较少的梯度上升轮次和较多的补偿训练可以有效减小遗忘数据对模型的干扰. 此外, 在补偿训练中, 仅需要 25% 的剩余用户参与, 这有助于减少计算开销, 同时保证补偿过程的高效性. 所有方法均采用存储历史更新的策略, 即剔除与遗忘用户相关的最新历史更新, 并在此基础上聚合其余用户的最新参数, 作为初始遗忘模型, 从而确保遗忘过程的有效性.

实验结果如图 2 所示, 比较了 FR, FE, RR, Shard 和 Ours 在独立同分布 (independent and identically distributed, IID) 和非独立同分布 (Non-IID) 条件下的 MNIST, Fashion-MNIST 和 CIFAR-10 数据集上的遗忘模型准确率. 结果表明, 在 IID 或 Non-IID 的情况下, FR (即重训练方法) 在准确率方面始终表现最佳. 这是因为重训练直接移除了目标用户的数据, 并对模型进行了完整的重新训练, 因此保证了模型在移除数据后能够恢复到接近原始模型的性能. 随后的是 Shard 和 Ours, 这两种方法通过不同的分片和校准策略有效地减少了遗忘过程对剩余数据的影响. 而 FE 和 RR 由于更多依赖于局部更新和遗忘操作, 无法像重训练方法那样全面恢复模型性能, 因此它们的准确率相对较低.

在遗忘时间方面, 表 1 显示了各方法的时间开销. FR 所需的遗忘时间最长, 至少是其他方法的两倍, 这是因为重训练涉及完整的数据重新训练, 需要更多的计算资源和时间. 相比之下, Ours 通过梯度上升和补偿训练策略, 最大限度地减少了全局重新训练的次数, 同时仅让部分用户参与补偿训练, 从而显著缩短了遗忘时间. 而 Shard 由于分片策略在一定程度上隔离了数据的相互影响, 也能较为有效地控制遗忘时间. FE 和 RR 在遗忘过程中更多地依赖于历史数据的校准更新, 尽管它们在一定程度上加快了遗忘速度, 但与 Ours 相比仍略显低效.

在基于成员推理攻击 (membership inference attack, MIA) 的 F1 score 方面, FR 重训练方法表现最优, 因为重训练几乎完全消除了对遗忘数据的推断能力. 为了对其他方法进行排名, 本文根据每种方法的 F1 score 与 FR 的差距来进行评价. 差值越小, 表示该方法在遗忘数据方面越接近重训练效果.

https://www.sciengine.com/doi/10.1360/SSI-2024-0304

表 1　单层语义知识库遗忘模型性能.

Table 1　Performance of the unlearned model for single-layer semantic knowledge base.

		MNIST		FMNIST		CIFAR-10	
		IID	Non-IID	IID	Non-IID	IID	Non-IID
F1 score	FL [32]	0.9556	0.9567	0.9468	0.9540	0.9525	0.9639
	FR [18]	0.5692	0.5724	0.5560	0.5240	0.5093	0.5381
	FE [18]	0.5438	0.3249	0.5289	0.3452	0.5543	0.4167
	RR [19]	0.5380	0.5990	0.5062	0.4486	0.5107	0.5044
	Shard [20]	0.5682	0.5167	0.5475	0.5975	0.5009	0.4290
	Ours	0.5446	0.5545	0.5436	0.5242	0.4621	0.4451
		IID	Non-IID	IID	Non-IID	IID	Non-IID
Unlearning time (s)	FR [18]	357.84	352.51	358.04	353.16	356.03	347.91
	FE [18]	176.92	180.03	175.92	177.37	186.30	195.83
	RR [19]	237.61	249.42	239.45	251.67	262.81	255.92
	Shard [20]	57.49	56.78	60.15	58.17	63.01	62.89
	Ours	57.70	57.96	55.26	56.68	61.60	60.28

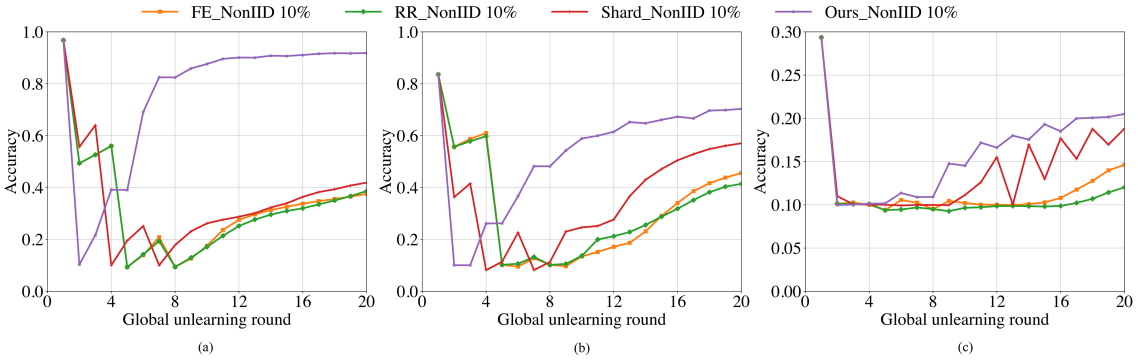


图 3　(网络版彩图) 在非独立同分布且存在 10% 恶意用户的情况下, 单层语义知识库遗忘模型的准确率. (a) MNIST; (b) Fashion-MNIST; (c) CIFAR-10.

Figure 3　(Color online) Accuracy for single-layer semantic knowledge base with 10% malicious users under Non-IID conditions. (a) MNIST; (b) Fashion-MNIST; (c) CIFAR-10.

因此, 差值依次为: Ours (0.1953), RR (0.2562), Shard (0.2181), FE (0.6452). 根据这些差值, F1 score 的排名依次为: FR, Ours, Shard, RR, FE. 这表明 Ours 在遗忘效果上最接近重训练方法, 其次是 RR 和 Shard, 而 FE 表现最差.

本文将准确率、遗忘时间和 F1 score 视为具有相同权重的核心指标, 由于每个指标的量纲不同, 将 3 个指标的排名相加, 得到每种方法在不同维度上的综合表现, 是一种简化的多维度评估方法. 综合每个方法在准确率、遗忘时间和 F1 score 中的排名, 最终得出各方法的总得分: Ours 得分为 6, FR 为 7, Shard 为 8, FE 和 RR 均为 12. Ours 方案之所以表现最佳, 是因为它在保证较高遗忘准确率和接近重训练的遗忘效果的同时, 显著减少了遗忘时间开销, 验证了高效性.

为验证提出方案的可信性, 实验假设在遗忘过程中存在恶意节点干扰. 在 Shard, FE 和 RR 这 3 种由聚合节点发起的方案中, 每轮聚合和共享过程中, 有 10% 或 20% 的概率聚合恶意参数. 而对于 Ours 方案 (由遗忘用户发起), 则在遗忘补偿训练过程中, 也有 10% 或 20% 的概率聚合恶意参数. 图 3 和 4 分别展示了在非独立同分布的 MNIST, Fashion-MNIST 和 CIFAR-10 数据集上的表现. 从实验结果可以看出, Ours 方案不受恶意节点的影响. 这是因为其采用了可信执行环境, 当检测到不一致时,

https://www.sciengine.com/doi/10.1360/SSI-2024-0304

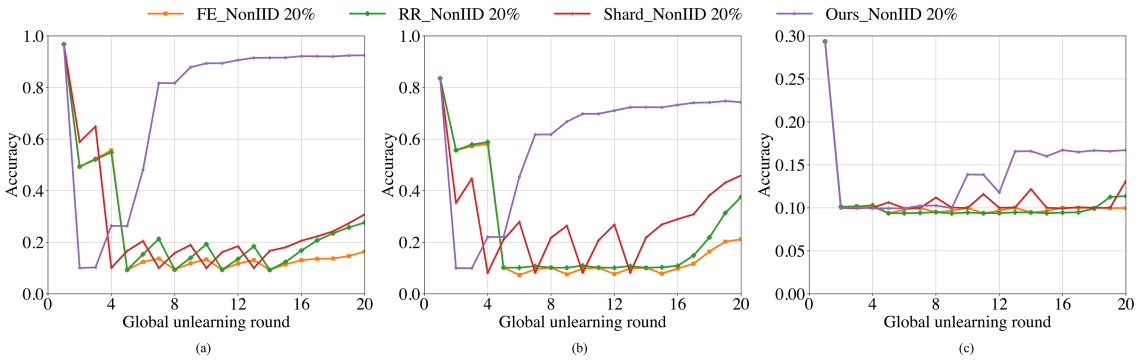


图 4　(网络版彩图) 在非独立同分布且存在 20% 恶意用户的情况下, 单层语义知识库遗忘模型的准确率. (a) MNIST; (b) Fashion-MNIST; (c) CIFAR-10.

Figure 4　(Color online) Accuracy for single-layer semantic knowledge base with 20% malicious users under Non-IID conditions. (a) MNIST; (b) Fashion-MNIST; (c) CIFAR-10.

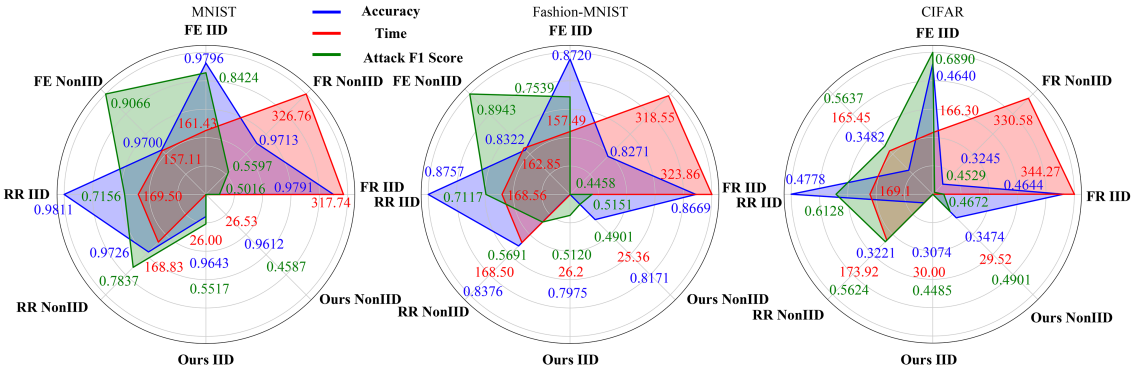


图 5　(网络版彩图) 多层语义知识库遗忘性能.

Figure 5　(Color online) Performance of the unlearned model for the multi-layer semantic knowledge base.

会返回至上一轮可信的结果并继续补偿训练, 确保模型的稳定性. Shard 方案虽然受到了恶意节点的干扰, 但影响较小. 其分片隔离机制使得某个分片中的恶意节点对全局的影响被分散, 恶意分片与其他正常分片聚合, 减少了恶意参数对整体模型的干扰. 相比之下, FE 和 RR 对恶意节点的干扰更为敏感. 由于这两种方法依赖于全局聚合, 恶意参数的加入会显著降低模型性能, 尤其是在较高概率恶意参数参与聚合的情况下, 模型的稳定性和可靠性会受到更大影响.

4.3　多层语义知识库遗忘实验结果

在多层语义知识库的遗忘方案中, 由于 Shard 方法不允许分片之间的结果聚合与共享, 因此本节仅讨论 FR, FE, RR 和 Ours 这 4 种方案. 值得注意的是, FE 和 RR 在提出时是针对单层架构设计的, 本文对其进行了适配, 以应用于多层架构. 在多层架构中, Ours 方案相较于单层架构有重要区别: 不再使用补偿训练, 而是采用了遗忘校准操作, 以更有效地减少遗忘用户对模型的长期影响. 这是因为随着多轮次的层层聚合, 遗忘用户的历史更新已经深度嵌入到模型中, 单纯依靠补偿训练难以完全消除其影响. 为了平衡效果和效率, 遗忘校准仅执行 10 个轮次, 其余实验配置保持不变.

实验结果如图 5 所示. 蓝色表示模型遗忘后的准确率, 红色表示遗忘所需的时间, 绿色表示基于成员推理攻击的 F1 score. 图中展示的是经过归一化处理后的实验数据, 用雷达图表示各方法的综合表现. 在遗忘后模型准确率方面, 排名依次为 RR, FE, FR, Ours. RR 和 FE 在准确率上表现较好, 表明它们在保留模型性能方面具备一定优势. 然而, 在遗忘时间方面, Ours 表现最优, 排名依次为 Ours, FE, RR, FR. 在基于成员推理攻击的 F1 score 排名中, FR 表现最佳, 之后依次为 Ours, RR, FE. 相较

https://www.sciengine.com/doi/10.1360/SSI-2024-0304

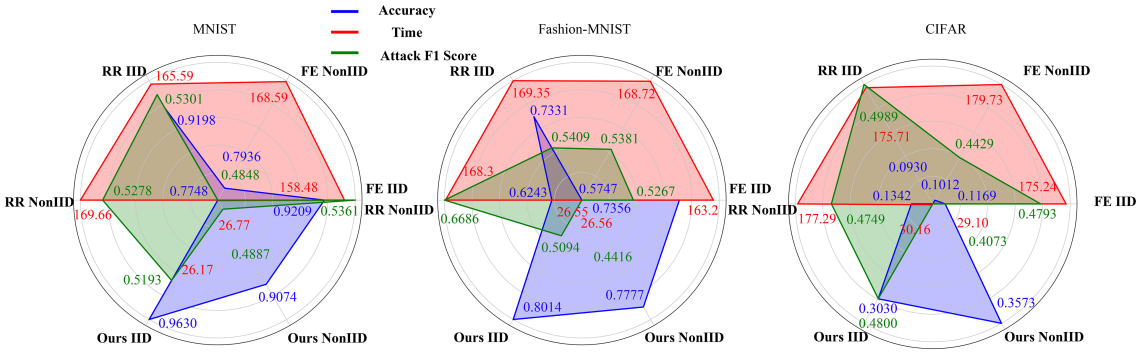


图 6　(网络版彩图) 存在 10% 恶意用户的情况下, 多层语义知识库遗忘模型性能.

Figure 6　(Color online) Performance for multi-layer semantic knowledge base with 10% malicious users.

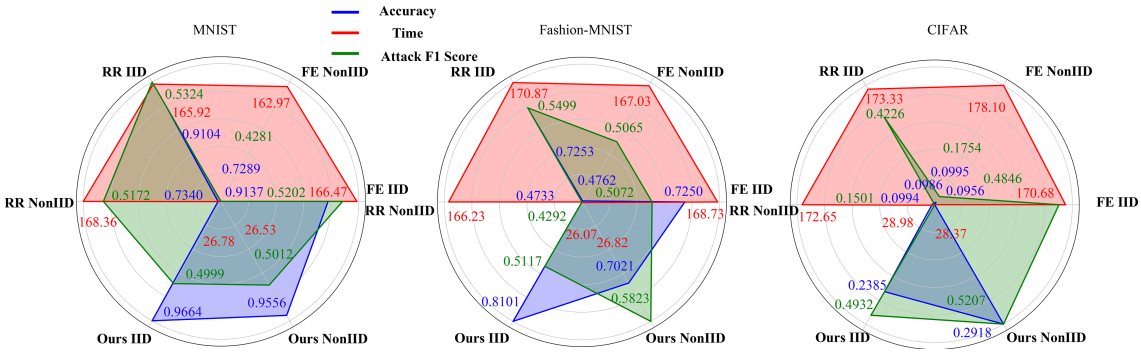


图 7　(网络版彩图) 存在 20% 恶意用户的情况下, 多层语义知识库遗忘模型的性能.

Figure 7　(Color online) Performance for multi-layer semantic knowledge base with 20% malicious users.

于单层遗忘, RR 和 FE 的 F1 score 未能显著降低, 表明这两种方法难以有效解决多层架构中无效数据的累积和放大影响. 综合 3 个指标的排名, Ours 和 RR 总得分为 7, 优于 FE 和 FR (总得分为 8). 这表明 Ours 在牺牲模型准确性的前提下, 显著降低了遗忘时间和开销, 并且能够有效应对多层架构中的数据累积问题. 相反, RR 虽然在准确率上表现突出, 但由于其高时间消耗和对累积数据影响的弱控制, 整体表现不如预期. 多层语义知识库架构下, 累积的数据影响对遗忘效果至关重要. 虽然 RR 和 FE 能够提供较高的准确率, 但在多层架构中, 其未能有效消除累积数据的影响. 而 Ours 方案通过遗忘校准机制, 成功在多层环境中消除无效数据的影响, 从而在效率和遗忘效果之间达到了较好的平衡.

为验证方案在多层架构中的可信性, 本实验采用了与单层架构相同的配置. 实验结果如图 6 和 7 所示. 在遗忘后模型准确率方面, Ours 几乎未受到恶意节点的影响, 表现出较强的鲁棒性. 相比之下, RR 和 FE 受恶意节点干扰显著, 准确率明显下降. 在 F1 score 方面, RR NonIID 和 FE NonIID 由于模型性能下降, F1 score 较低, 未能充分反映出有效的遗忘性能. Ours 通过遗忘校准机制, 能够在 Non-IID 条件下有效应对多层架构中的数据累积问题, 从而保持了较高的 F1 score. 在遗忘时间方面, 结果与不存在恶意节点的情况一致. Ours 在控制时间开销方面依旧表现最优. 综上所述, Ours 在多层架构下依旧表现出较强的抗恶意节点能力, 同时有效保障了模型的准确率和遗忘效果. 相反, RR 和 FE 在恶意节点存在的情况下, 难以维持模型性能, 且在 Non-IID 条件下更为明显.

5　结论

本文面向多用户语义通信场景提出语义知识库高效可信联邦遗忘解决方案, 旨在无需重新训练的前提下消除无效数据影响并维持语义知识库的可用性. 首先, 构建了基于联邦遗忘的多用户语义知识

https://www.sciengine.com/doi/10.1360/SSI-2024-0304

库更新框架, 以支持在多用户语义通信场景下消除无效数据影响; 其次, 提出了基于补偿梯度上升和可信执行环境的单层语义知识库遗忘方法, 确保单层多用户语义通信场景下高效可信地遗忘无效数据; 同时, 提出了基于梯度上升和历史梯度的多层语义知识库遗忘方法, 以应对跨域多层级语义通信中无效数据的复杂影响, 从而通过历史梯度信息的整合确保全局知识库的一致性和准确性; 最后, 仿真结果从遗忘时间、遗忘有效性和遗忘后语义知识库可用性 3 个方面验证提出方案. 本文关注的多用户协作环境中语义知识库消除无效数据影响问题, 具有重要的应用价值. 未来工作可以针对跨领域、跨层次的多用户语义通信, 进一步研究如何在异构设备和网络条件下提升语义知识库遗忘方案的适应性和扩展性, 使得该方案能够更好地应对复杂的应用场景和多用户协作需求.

参考文献

1 Reinsel D, Wu L, Gantz J F, et al. The China datasphere: primed to be the largest datasphere by 2025. In: Proceedings of International Data Cooperation (IDC), Framingham, 2019. 1–10
2 Yang S, Fang C, Zhong L, et al. A novel distributed multi-source optimal rate control solution for HTTP live video streaming. IEEE Trans Broadcast, 2024, 70: 792–807
3 Zhang P, Niu K, Yao S S, et al. Semantic communications for future: basic principle and implementation methodology. J Commun, 2023, 44: 1–14 [张平, 牛凯, 姚圣时, 等. 面向未来的语义通信: 基本原理与实现方法. 通信学报, 2023, 44: 1–14]
4 Sun Y P, Cui S G, Zhang P. Survey on semantic knowledge base for semantic communications. ZTE Commun, 2023, 29: 19–23 [孙亚萍, 崔曙光, 张平. 面向语义通信的语义知识库综述. 中兴通讯技术, 2023, 29: 19–23]
5 鹏城实验室, 北京邮电大学和香港中文大学 (深圳). 面向现代语义通信的语义知识库技术白皮书. 鹏城实验室, 2023
6 Yang Z, Chen M, Wong K K, et al. Federated learning for 6G: applications, challenges, and opportunities. Engineering, 2022, 8: 33–41
7 Tu Y F, Chen W. A deployment method of multi-user semantic communication system based on federated learning. J Signal Process, 2022, 38: 2486–2495 [涂勇峰, 陈文. 基于联邦学习的多用户语义通信系统部署方法. 信号处理, 2022, 38: 2486–2495]
8 Niu K, Dai J, Yao S, et al. A paradigm shift toward semantic communications. IEEE Commun Mag, 2022, 60: 113–119
9 Ren J, Zhang Z, Xu J, et al. Knowledge base enabled semantic communication: a generative perspective. IEEE Wireless Commun, 2024, 31: 14–22
10 Nguyen L X, Le H Q, Tun Y L, et al. An efficient federated learning framework for training semantic communication systems. IEEE Trans Veh Technol, 2024, 73: 15872–15877
11 Lu X L, Zhu K, Li J, et al. Efficient knowledge base synchronization in semantic communication network: a federated distillation approach. In: Proceedings of IEEE Wireless Communications and Networking Conference (WCNC), 2024. 1–6
12 Nagalapatti L, Mittal R S, Narayanam R. Is your data relevant? Dynamic selection of relevant data for federated learning. In: Proceedings of the AAAI Conference on Artificial Intelligence, 2022. 7859–7867
13 Li Q M, Li X H, Zhou L, et al. AdaFL: adaptive client selection and dynamic contribution evaluation for efficient federated learning. In: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2024. 6645–6649
14 Deng Y, Lyu F, Ren J, et al. AUCTION: automated and quality-aware client selection framework for efficient federated learning. IEEE Trans Parallel Distrib Syst, 2022, 33: 1996–2009
15 Blanchard P, Mhamdi E M E, Guerraoui R, et al. Machine learning with adversaries: Byzantine tolerant gradient descent. In: Proceedings of Advances in Neural Information Processing Systems, 2017
16 Xia Q, Tao Z Y, Tao Z J, et al. FABA: an algorithm for fast aggregation against Byzantine attacks in distributed neural networks. In: Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI), 2019
17 Zhao L, Jiang J, Feng B, et al. SEAR: secure and efficient aggregation for Byzantine-robust federated learning. IEEE Trans Dependable Secure Comput, 2022, 19: 3329–3342
18 Liu G Y, Ma X Q, Yang Y, et al. FedEraser: enabling efficient client-level data removal from federated learning models. In: Proceedings of IEEE/ACM 29th International Symposium on Quality of Service (IWQOS), 2021. 1–10
19 Liu Y, Xu L, Yuan X L, et al. The right to be forgotten in federated learning: an efficient realization with rapid retraining. In: Proceedings of IEEE Conference on Computer Communications, 2022. 1749–1758
20 Lin Y J, Gao Z P, Du H Y, et al. Scalable federated unlearning via isolated and coded sharding. In: Proceedings of the 33rd International Joint Conference on Artificial Intelligence, 2024. 4551–4559
21 Su N X, Li B C. Asynchronous federated unlearning. In: Proceedings of IEEE Conference on Computer Communications, 2023. 1–10
22 Lin Y, Du H, Gao Z, et al. Decentralized unlearning for trustworthy AI-generated content (AIGC) services. IEEE Netw, 2024. doi: 10.1109/MNET.2024.3439411
23 Ding N N, Sun Z Y, Wei E, et al. Incentive mechanism design for federated learning and unlearning. In: Proceedings of the 24th International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing, 2023. 11–20
24 Lin Y J, Gao Z P, Du H Y, et al. Incentive and dynamic client selection for federated unlearning. In: Proceedings of

https://www.sciengine.com/doi/10.1360/SSI-2024-0304

the ACM on Web Conference, 2024. 2936–2944
25 Bourtoule L, Chandrasekaran V, Choquette-Choo C, et al. Machine unlearning. In: Proceedings of IEEE Symposium on Security and Privacy (SP), 2021. 141–159
26 Li X, Huang K X, Yang W H, et al. On the convergence of FedAvg on non-IID data. In: Proceedings of International Conference on Learning Representations, 2020
27 Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition. Proc IEEE, 1998, 86: 2278–2324
28 Xiao H, Rasul K, Vollgraf R. Fashion-MNIST: a novel image dataset for benchmarking machine learning algorithms. 2017. ArXiv:1708.07747
29 Krizhevsky A, Hinton G. Learning multiple layers of features from tiny images. 2009. http://www.cs.utoronto.ca/kriz/learning-features-2009-TR.pdf
30 Wang H, Kaplan Z, Niu D, et al. Optimizing federated learning on non-IID data with reinforcement learning. In: Proceedings of IEEE Conference on Computer Communications, 2020. 1698–1707
31 Shokri R, Stronati M, Song C, et al. Membership inference attacks against machine learning models. In: Proceedings of IEEE Symposium on Security and Privacy (SP), 2017. 3–18
32 McMahan B, Moore E, Ramage D, et al. Communication-efficient learning of deep networks from decentralized data. In: Proceedings of Artificial Intelligence and Statistics, 2017. 1273–1282

Efficient and trusted federated unlearning for multi-user semantic knowledge base

Yijing LIN[1], Jinke REN[2,4*], Jiacheng WANG[5*], Nan MA[1,3], Ye ZHANG[6] & Shuguang CUI[4,2]

1. *State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China*
2. *Shenzhen Future Network of Intelligence Institute, Shenzhen 518172, China*
3. *Department of Broadband Communication, Peng Cheng Laboratory, Shenzhen 518055, China*
4. *School of Science and Engineering and Guangdong Provincial Key Laboratory of Future Networks of Intelligence, The Chinese University of Hong Kong (Shenzhen), Shenzhen 518172, China*
5. *College of Computing and Data Science, Nanyang Technological University, Singapore 639798, Singapore*
6. *School of Computer Science, Beijing Information Science & Technology University, Beijing 100192, China*
* Corresponding author. E-mail: jinkeren@cuhk.edu.cn, jiacheng.wang@ntu.edu.sg

Abstract Semantic communication extracts useful semantic information from raw data by using a shared semantic knowledge base between transmitters and receivers. This approach reduces data volume and improves transmission efficiency. In multi-user semantic communications, the semantic knowledge base relies on data from multiple users to ensure accurate semantic extraction, a process facilitated by federated learning, which enables collaboration without sharing raw data. However, outdated, irrelevant, or harmful data may compromise the reliability of the semantic knowledge base. To address this challenge, this study proposes an efficient and trusted federated unlearning mechanism for multi-user semantic knowledge bases. The mechanism considers single-layer and multi-layer multi-user semantic communications and introduces a federated unlearning method based on gradient ascent and trusted execution environments; this method effectively eliminates the impact of invalid data while preserving reliability. Experimental results demonstrate that the proposed mechanism significantly reduces computational overhead while maintaining comparable accuracy in unlearned data and the post-unlearning semantic knowledge base.

Keywords multi-user semantic communication, semantic knowledge base, federated unlearning

https://www.sciengine.com/doi/10.1360/SSI-2024-0304