第 62 卷　第 5 期
2026 年 9 月

南京大学学报(自然科学)
JOURNAL OF NANJING UNIVERSITY
(NATURAL SCIENCE)

Vol. 62, No. 5
Sept., 2026

DOI:10.13232/j.cnki.jnju.2026.05.008

# 面向异质扰动场景的多模态联邦自适应关系蒸馏方法 AMRD-FedA$^2$

王一璇[1],杨　飞[2],许春根[2*]

(1.南京理工大学网络空间安全学院,南京,214443;2.南京理工大学数学与统计学院,南京,210094)

**摘　要**:针对多模态联邦学习中客户端模态缺失、标签缺失和数据高度异质导致的模型性能下降问题,提出结合关系蒸馏的AMRD-FedA$^2$方法. AMRD (Adaptive Multimodal Relation Distillation)通过评估不同模态的表征质量,自适应选择优势模态,并利用样本关系矩阵在模态间进行知识蒸馏,来缓解模态缺失和标签缺失带来的负面影响;FedA$^2$ (Federated Algorithm Adaptation Framework)将该蒸馏机制适配到不同联邦学习框架中,以增强方法在联邦优化场景下的通用性. 在传感器、文本、图像、视频和音频等模态数据集上以不同主流联邦学习算法为对比基线进行实验,结果表明在高度模态缺失和高度标签缺失场景下,框架的整体效果优于原始基线. 其中,以SCAFFOLD算法测试的Crisis-MMD数据集在实验条件下的$F1$分数提升幅度超过30%. 进一步研究发现,提出的方法在强non-IID (Non-Independent and Identically Distributed)场景下表现更优,在复杂高异质多模态联邦环境中具有较好的鲁棒性和泛化能力.

**关键词**:多模态联邦学习,模态缺失,标签缺失,关系蒸馏

**中图分类号**:TP181,TP391　　**文献标志码**:A

# AMRD-FedA$^2$: Adaptive relation distillation for multimodal federated learning under heterogeneous perturbations

Wang Yixuan[1], Yang Fei[2], Xu Chungen[2*]

(1. School of Cyberspace Security, Nanjing University of Science and Technology, Nanjing, 214443, China;
2. School of Mathematics and Statistics, Nanjing University of Science and Technology, Nanjing, 210094, China)

**Abstract**: To address the degradation in model performance caused by missing modalities, missing labels, and high data heterogeneity in multimodal federated learning, we propose AMRD-FedA$^2$, a method that integrates relational distillation. AMRD (Adaptive Multimodal Relation Distillation) assesses the representational quality of different modalities, adaptively selects the most informative modalities, and leverages sample-level relation matrices to perform knowledge distillation across modalities, thereby mitigating the adverse effects of modality and label. FedA$^2$ (Federated Algorithm Adaptation Framework) adapts this distillation mechanism to various federated learning frameworks, enhancing the method's versatility in federated optimization settings. Experiments conducted on multimodal datasets, including sensor, text, image, video, and audio, using several mainstream federated learning algorithms as baselines demonstrate that, under conditions of severe modality and label, our framework outperforms the original baselines. Notably, on the Crisis-MMD dataset evaluated with the SCAFFOLD algorithm, the $F1$ score improvement exceeds 30% under the experimental conditions. Further analysis reveals that our approach performs particularly well in strongly non-IID (Non-Independent and Identically Distributed) scenarios, exhibiting robustness and strong generalization capabilities in complex, highly heterogeneous multimodal federated environments.

**Keywords**: multimodal federated learning, missing modalities, missing labels, relation distillation

基金项目:江苏省研究生科研与实践创新计划(KYCX25_0817)

收稿日期:2026－07－14

*通信联系人,E-mail:xuchung@njust.edu.cn

随着大数据、物联网以及边缘智能的发展,数据采集和模型训练正在从中心化平台逐渐延伸到分布式终端设备.手机、可穿戴设备、车载终端、便携医疗设备和边缘物联网设备在使用过程中会持续生成传感器、音频、图像、视频和文本等数据,这些数据具有规模大、来源分散、隐私敏感和模态多样等特点,使传统集中式训练在数据传输、隐私保护和跨主体协作等方面受制.联邦学习的出现为隐私安全分布式数据建模提供了方案.Kairouz et al指出[1],统计异质性、系统异质性、隐私风险和通信效率是联邦学习长期存在的问题.高莹等[2]从攻击与防御角度说明,联邦学习虽然避免了原始数据集中,但模型更新过程仍可能引入隐私泄露与安全风险.大数据环境下的联邦学习不是简单的分布式训练框架,更需要满足日益增长的数据规模和隐私安全需求.

在多模态任务中,上述问题进一步复杂化.与传统单模态联邦学习相比,多模态联邦学习不仅要处理客户端数据分布不一致,还要面对不同客户端模态组合不一致、模态缺失和标签缺失等问题.Che et al[3]认为多模态联邦学习的关键挑战是在不共享原始数据的前提下实现跨模态信息协同.从系统实现角度看,林伟伟等[4]对联邦学习开源框架的分析表明,联邦算法和模型训练策略会直接影响联邦学习方法的可扩展性.这对多模态联邦学习具有启发意义,当模态类型、客户端能力和数据分布同时变化时,算法不能和单模态联邦学习一样只关注单一模型结构,还必须具备跨任务、跨模态和跨优化机制的适配能力.现有多模态联邦学习研究已经从不同角度尝试解决此问题,不再认为这是简单叠加多个模态,而是思考不同模态应该如何协同,思考不同结构的客户端如何共享知识.但是现有研究更多关注全局层面的模态协同或服务器端知识聚合,少有关注客户端本地模态表征的.

多模态模型训练过程中,当某些模态出现缺失、噪声或样本分布偏移时,模型可能主要依赖少数强模态完成训练,弱势模态的特征结构往往被忽略,当客户端数据分布同时也满足非独立同分布(non-IID)时更加剧这一点于改善弱势模态,Le et al[5]的FedMEKT利用知识迁移缓解训练中标签不足的问题,Wang et al[6]的FedMMR通过缺失模态重建改善模型性能.表明,模态知识补偿确实能改善多模态联邦学习中的信息不足问题,但在客户端本地训练过程中,动态判断当前哪个模态更可靠以及弱势模态应学习优势模态的何种信息,值得进一步研究.

模态之间的学习依赖特征值,但异构模态之间可能不适合直接对齐特征值,对于敏感度高的模态,任何微小扰动都可能导致语义的完全改变,所以更好的办法是对齐样本关系结构.本文将该问题转化为本地跨模态关系蒸馏,即先选择可靠模态,再利用样本关系结构引导弱势模态优化,用样本关系矩阵作为跨模态蒸馏媒介,避免不同模态特征空间不一致带来的直接对齐冲突.

针对上述现有方法的局限性,本文提出一种同时处理数据异构、模态异构和客户端本地不同模态表征差异的方法.该方法在客户端本地评估不同模态的可靠性,动态选择优势模态,并通过样本关系矩阵引导弱势模态进行结构学习,缓解因模态缺失、标签缺失和高度non-IID导致的模态表征失衡.

本文的主要贡献如下.

(1)提出一种自适应多模态关系蒸馏算法(Adaptive Multimodal Relation Distillation, AMRD),通过模态可靠性评分动态识别客户端本地优势模态,并利用样本关系矩阵引导弱势模态学习优势模态的结构信息,可以缓解因模态缺失、标签缺失和高度non-IID导致的模态表征失衡问题.

(2)提出一种联邦算法适配框架(Federated Algorithm Adaptation Framework, FedA$^2$),将常见的联邦学习算法差异抽象为统一的本地算法适配项,使AMRD能够在不重构联邦训练流程的前提下适配不同联邦优化机制.

(3)构造多模态数据集的通用集合表示方法,将传感器、音频、视频、图像和文本等不同类型模态统一表示为可比较、可蒸馏的模态集合,使数据集能够在不同模态的自由组合下完成可靠模态选择与关系蒸馏.

## 1 相关工作

**1.1 多模态联邦学习** 多模态联邦学习研究的核心是当数据分散在不同客户端且每个客户端拥有的模态类型、模态数量和数据分布不一致时,如何实现较好的联合建模. 与传统单模态联邦学习相比,多模态联邦学习的难点不只来自客户端数据分布差异,还来自模态间语义空间、采样机制和特征结构的不一致. Che et al[3]指出多模态联邦学习需要同时处理模态融合、模态缺失、模型异构和隐私保护等问题,这一判断为本文提供了基本出发点,若只在融合层之后优化分类损失,模型可能忽略各模态编码器之间的质量差异,因此需要在本地训练阶段显式确认模态间的关系.

围绕模态组合不一致问题,Chen and Zhang[7]的FedMSplit将不同客户端拥有的传感器组合为多模态分裂网络,通过自适应多任务学习来缓解模态组合不一致的问题. 该方法虽然强调客户端模态结构不一致会影响联邦训练,但其主要关注的是网络结构层面. Ouyang et al[8]的Harmony将多模态训练拆分为模态级联邦学习和融合学习两个阶段,来降低异构客户端之间的训练冲突. 该方法确实降低了不同模态客户端之间直接聚合引发的冲突,但更强调训练流程分解.

跨模态知识传递进一步推动了多模态联邦学习的发展,如Yu et al[9]的CreamFL通过对比表示来集成实现异构客户端之间的多模态知识迁移. 说明多模态联邦学习不应局限于参数平均,还可以通过表示知识或嵌入知识传递实现协作. 然而现有方法大多强调服务端全局知识共享或客户端间的知识传递,有关客户端本地强模态如何指导弱模态学习的讨论仍不足.

多模态联邦学习研究表明,客户端数据异构与模态结构异构会共同影响模型的训练,两者不是独立影响. 而且,随着异构程度的不同,影响的强度也有变化. 后续应该对异构类型进行区分研究.

**1.2 模态缺失、标签缺失与错误标签** 单模态联邦学习中不存在模态缺失,但是在多模态联邦学习中,这是一种最直接也最有特点的扰动. 在实际场景中,传感器损坏、采集设备的差异、网络传输异常、音轨丢失、图像损坏和文本内容缺失,都可能导致客户端只具备部分模态信息. 在同一客户端的内部,不同样本的模态可用性也可能不同. Le et al[10]的跨模态原型学习方法在严重模态缺失场景下利用模态特异原型改善模型性能. 这种方法试图补偿缺失模态本身,但缺失模态重建通常需要额外的强假设条件,过于理想化,不适合本文对通用方法的研究.

标签缺失同样会削弱训练过程中的本地约束信号. 在数据整理工作中,人工标注的成本高,面对复杂数据类型时难以获得完整标签,用这样的数据进行训练会导致常规分类损失只能作用于有限样本,会大幅降低训练效果,且没有标签的数据也会被浪费. FedMEKT方法[5]说明知识迁移可以为本地模型提供额外约束,来缓解标签缺失的部分影响. 对于数据分布,标签缺失还可能与类别不平衡同时出现,使部分类别和弱势模态长期处于监督不足的状态. 此外,客户端类别分布不均也会影响局部判别边界和全局模型泛化能力[11],这都表明弱模态问题不一定只由模态缺失造成,也可能由标签稀缺和客户端分布偏移共同导致.

错误标签会直接污染监督目标,使模型学习到错误的类别关系. 与模态缺失和标签缺失相比,错误标签不仅减少了信息量,还引入了错误方向的优化信号.

综上,模态缺失、标签缺失和错误标签虽然表现不同,但都会进一步加剧本地模态表征质量不均衡,最终都会影响客户端本地模态表征质量.

**1.3 知识蒸馏与关系蒸馏** 知识蒸馏为多模态联邦学习提供了比参数聚合更灵活的知识迁移方式. Gou et al[12]将蒸馏知识划分为响应知识、特征知识和关系知识等类型. 潘海为等[13]从视觉算法角度总结了知识蒸馏的发展,指出关系蒸馏能够在不强制特征值完全一致的情况下传递结构信息. 这对多模态联邦学习尤其重要,因为不同模态的特征空间可能存在天然差异,直接对齐特征向量容易破坏模态特异性.

在关系蒸馏方向,Park et al[14]的Relational KD将样本集合之间的距离和角度关系作为知识进行迁移. Tung and Mori[15]的Similarity-Preser-

ving KD使学生模型保持教师模型中的样本相似性结构. 表明教师模型真正有价值的信息不一定只是单个样本的类别输出，也可以是样本集合内部的结构关系. 本文提出的AMRD借鉴这一思想，将每个模态的batch特征转换为样本关系矩阵，并让弱势模态学习优势模态中的样本相似结构，既避免了异构模态直接特征对齐，也使传感器、音频、图像、视频和文本等不同模态可以被统一纳入关系蒸馏框架.

在多模态联邦场景中，关系蒸馏需要确定教师模态. Guo et al[16]的MASA在多模态联邦学习中通过模态感知机制和门控跨模态蒸馏实现优势模态向弱势模态的信息流动，说明模态间的知识迁移应当具有方向性，而不是简单的双向拉近. 本文提出的AMRD进一步将这一思想抽象为通用本地优化机制，通过模态可靠性分数动态确定教师模态，通过分数阈值避免错误蒸馏，通过分类损失上界控制蒸馏强度. 由此，AMRD不仅是跨模态蒸馏方法，也是一种面向本地模态表征失衡的结构正则化方法.

知识蒸馏与关系蒸馏说明异构模态之间可以通过样本关系结构传递知识，这为复杂多变的多模态数据之间搭起了桥梁，尤其在面对异质性较强的不同模态数据时，这种跨模态蒸馏方式有天然的优势.

**1.4　联邦优化算法**　联邦优化算法决定了客户端本地目标和模型更新方式. McMahan et al[17]的FedAvg通过客户端本地多步训练和服务器加权平均来降低通信频次，是多数联邦算法的基础. FedAvg的优势是简单高效，但在non-IID条件下容易出现客户端更新方向不一致. Li et al[18]的FedProx在本地目标中加入近端项限制本地模型偏离全局模型. Karimireddy et al[19]的SCAFFOLD通过控制变量来修正客户端更新方向，减轻客户端漂移. 这两类方法说明，联邦优化的关键并不是简单更新参数，而是如何控制本地训练与全局目标之间的偏移.

除标签分布偏移外，特征分布偏移也会影响联邦训练. Li et al[20]的FedBN将Batch Normalization相关参数保留在客户端本地，以缓解不同客户端特征分布差异. Reddi et al的FedOpt[21]将客户端更新输入服务器端自适应优化器，提高了联邦优化在异构环境下的稳定性. 这些算法解决问题的角度不同：FedProx强调距离约束，SCAFFOLD强调方向修正，FedBN强调参数作用范围，FedOpt强调更新使用方式. 若AMRD分别为每种算法单独设计损失函数或输出规则，会导致方法复杂度高且不利于扩展，故AMRD应该始终负责输出跨模态关系蒸馏损失，FedA²负责将具体联邦算法转化为统一的本地适配约束，两者解耦互不干扰. 这样既能处理客户端内部的模态表征失衡，也能适配不同联邦优化机制，为后续扩展更多联邦算法提供统一接口.

联邦优化算法研究表明，不同算法对本地目标、参数范围和更新方向的要求不同. 现有研究虽然分别从模态异构、缺失补偿、知识迁移和联邦优化角度取得进展，但仍缺少一种同时面向本地模态结构补偿和多联邦算法适配的方法.

## 2　本文方法

**2.1　整体框架**　设任意客户端拥有任意模态种类数据. 在单轮本地训练任务中，会先进行模态编码和特征提取，模态数量满足条件时会开启AMRD算法. 随后，根据参数进行共同样本筛选，生成的样本关系矩阵也受适配算子约束，蒸馏后得到蒸馏损失项. 同时，FedA²根据联邦算法确定参数生效范围，生成算法适配项. 最后，合并损失为本地目标函数，进行本地模型更新，并上传服务器参与聚合. 整体框架如图1所示.

**2.2　AMRD优化目标**　联邦学习流程中应先由客户端完成本地训练，再上传模型更新. AMRD的优化位置位于客户端本地训练阶段，在传统联邦分类损失之外额外加入跨模态关系约束. 因为所有联邦学习算法都希望找到一组全局模型参数$W$，让所有客户端本地损失$F(W)$的加权和最小，因此在未引入AMRD时，多模态联邦学习目标均可以如式(1)所示. 式(1)先定义客户端本地分类风险，再通过客户端权重$p_k$得到全局目标$F(W)$. 本文常用的符号解释如表1所示.

$$\min_{W} F(W)=\sum_{k\in K} p_k \frac{1}{\left|D_k^l\right|} L_{\mathrm{cls}}\left(f_W\left(X_{k,i},M_k\right),y_{k,i}\right) \tag{1}$$

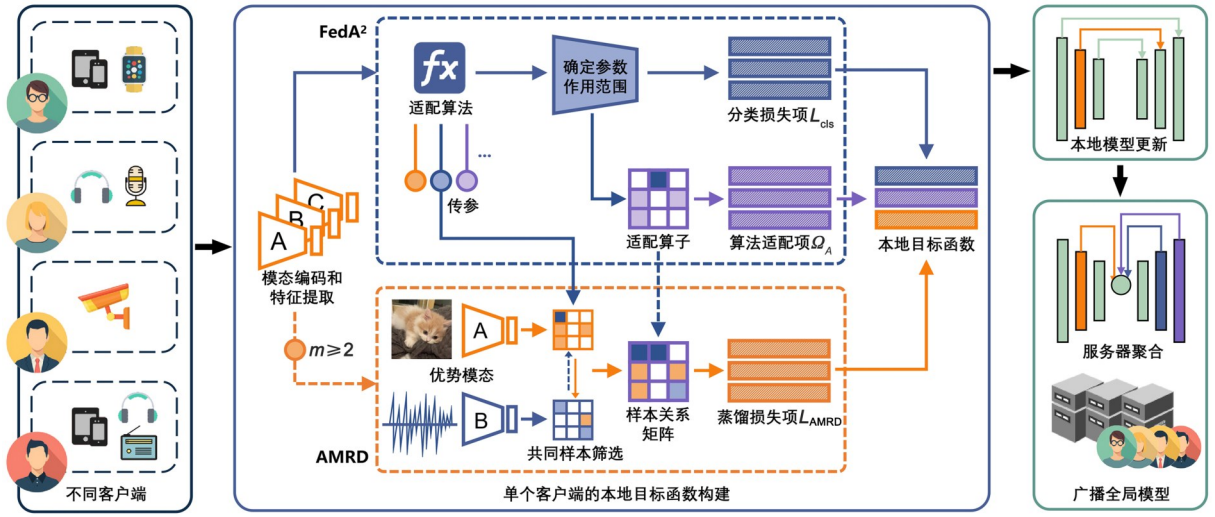



**图 1　整体框架结构图**

**Fig. 1　Overview of framework structure diagram**

**表 1　符号定义**

**Table 1　Definition of Formula Symbols**

| 符号 | 含义 | 符号 | 含义 |
|---|---|---|---|
| $K$ | 客户端集合,$K=\{1,2,\cdots,k\}$ | $\tilde{Y}_{k,t}^{m}$ | 聚类标签集合 |
| $k$ | 客户端索引 | $I(;)$ | 集合之间的互信息 |
| $M$ | 全局模态集合,$M=\{1,2,\cdots,m\}$ | $H(\cdot)$ | 标签集合的信息熵 |
| $m$ | 系统中可能出现的模态数量 | $m_k^+$ | 客户端 $k$ 当前选择的教师模态 |
| $M_k$ | 客户端 $k$ 可访问的模态子集,$M_k \subseteq M$ | $S_k$ | 客户端 $k$ 当前参与蒸馏的学生模态集合 |
| $X_{k,i}$ | 客户端 $k$ 的第 $i$ 个多模态样本整体输入 | $z_{k,i}^{m}$ | 客户端 $k$ 中的样本 $i$ 在模态 $m$ 下的特征表示 |
| $W$ | 多模态模型全局参数 | $\boldsymbol{R}_{k,t}^{m}$ | 客户端 $k$ 在 $t$ 轮中模态 $m$ 的样本关系矩阵 |
| $f_W$ | 参数为 $W$ 的多模态模型 | $\boldsymbol{R}_{k,t}^{+}$ | 教师模态 $m_k^+$ 在 $t$ 轮的关系矩阵 |
| $D_k^l$ | 客户端 $k$ 的第 $l$ 个本地标签样本 | $L_{cls}$ | 联邦算法分类损失函数 |
| $s_{k,t}^{m}$ | 客户端 $k$ 上模态 $m$ 在 $t$ 轮的可靠性分数 | $L_{rel}$ | 跨模态关系蒸馏损失函数 |
| $\bar{s}_{k,t}^{m}$ | 平滑后的可靠性分数 | $L_{AMRD}$ | AMRD 自适应多模态蒸馏损失函数 |
| $Y_{k,t}^{m}$ | 客户端 $k$ 上模态 $m$ 在 $t$ 轮的真实标签集合 | $J_A$ | 结合联邦算法 $A$ 后的客户端本地目标 |

其中,$p_k$ 通常由客户端样本量决定,$L_{cls}$ 表示分类损失,在联邦算法中通常为交叉熵损失.

当不同客户端存在模态缺失、标签缺失或数据分布差异时,仅依靠分类损失难以保证弱势模态学习到稳定表征结构.因此,AMRD 算法在此基础上加入本地跨模态蒸馏项,式(2)为扩展后的多模态联邦学习优化目标:

$$\min_{W} F_{AMRD}(W)=\sum_{k\in K} p_k\left[F_k(W)+\frac{1}{|T_k|}\sum_{t\in T_k} L_{AMRD}^{k,t}(W)\right] \tag{2}$$

其中,$F_k(W)$ 保证模型对标签的判别能力,$L_{AMRD}^{k,t}(W)$ 是自适应多模态蒸馏损失平均值,用来约束弱势模态学习优势模态的样本关系结构.

与式(1)相比,AMRD 在每个客户端的基础监督目标 $F_k(W)$ 上增加了跨模态关系蒸馏项.优化重点在客户端本地训练阶段,由于引入了额外约束,使弱势模态在保持原本分类监督的同时去学习优势模态的样本关系信息.

**2.3　模态可靠性选择与关系蒸馏**　在多模态联邦学习任务中,模型可能主要依赖强模态完成分

类和预测，弱模态的特征表现始终较差，需要先判断当前客户端中哪个模态更可靠，再让弱模态学习强模态中的结构信息.

设客户端 $k$ 在第 $t$ 轮中可用于可靠性评估的模态集合为 $E_{k,t}$. 对任意模态 $m$，AMRD根据该模态特征聚类标签与真实标签之间的一致性计算互信息和可靠性分数[16]：

$$s_{k,t}^{m}=\frac{2I\left(Y_{k,t}^{m};\tilde{Y}_{k,t}^{m}\right)}{H\left(Y_{k,t}^{m}\right)+H\left(\tilde{Y}_{k,t}^{m}\right)+\varepsilon} \tag{3}$$

式(3)中，$s_{k,t}^{m}$ 越大，表示模态 $m$ 在客户端 $k$ 的当前特征空间越能形成与类别标签一致的结构，此分数会随客户端、通信轮次和扰动状态的变化而变化. 为了降低单个轮次带来的偶然波动，实际选择教师模态时使用平滑后的分数 $\bar{s}_{k,t}^{m}$. 基于可靠性分数，AMRD可以选择当前最可靠的模态作为教师模态，并让分数差达到阈值的弱势模态参与蒸馏：

$$S_{k,t}=\left\{m\,\middle|\,\min E_{k,t}\left\{m_{k,t}^{+},\bar{s}_{k,t}^{m_{k,t}^{+}}-\bar{s}_{k,t}^{m}\geqslant\delta\right\}\right\} \tag{4}$$

式(4)中，不同客户端可以选择不同教师模态，同一客户端在不同训练阶段也可以切换教师模态. 阈值 $\delta$ 的作用是过滤不明显的模态差异，避免在教师优势不足时强行蒸馏.

确定教师和学生后，AMRD不能直接对齐两个模态的特征值. 由于不同模态数据集内的特征空间具有截然不同的维度和语义，直接拉近特征向量会破坏模态特异性，所以只能对齐样本之间的关系结构. 对于模态 $m$，其样本关系矩阵定义如式(5)所示，该矩阵描述同一batch内样本之间的相似关系：

$$\begin{aligned}&\boldsymbol{R}_{k,t}^{m}=\frac{\Phi\left(Z_{k,t}^{m}\right)\Phi\left(Z_{k,t}^{m}\right)^{\mathrm{T}}}{\tau}\\&\Phi\left(Z_{k,t}^{m}\right)=Norm_{\mathrm{row}}\left(H_{n}Z_{k,t}^{m}\right)\end{aligned} \tag{5}$$

其中，$\Phi(\cdot)$ 表示中心化和归一化后的特征变换，$\tau$ 为温度系数，$Norm_{\mathrm{row}}(\cdot)$ 表示按行进行 $L_2$ 归一化. 对于任意学生模态 $m$，AMRD使其关系矩阵靠近教师模态关系矩阵：

$$L_{\mathrm{rel}}^{k,t,m}=\frac{1}{n^{2}}\left\|\boldsymbol{R}_{k,t}^{m}-sg\left(\boldsymbol{R}_{k,t}^{+}\right)\right\|_{F}^{2} \tag{6}$$

其中，$sg(\cdot)$ 表示停止梯度，可以保证知识只从教师模态流向学生模态，避免弱模态反向干扰优势模态. 通过检测模态之间的差异，可以动态进行模态间的样本关系蒸馏.

**2.4　AMRD本地目标函数**　2.3给出算法的局部约束结构，本节给出具体的本地目标函数. 由于不同客户端和不同模态之间的差异始终在变化，不应该在所有样本和所有轮次中使用固定蒸馏强度. 本地目标函数应该同时考虑三个因素，即教师模态优势是否明显、训练阶段是否稳定、蒸馏项是否过强.

首先，根据式(4)得到的学生模态集合 $S_{k,t}$，对每个学生模态计算关系蒸馏损失. 随后，利用每一轮的权重 $\omega_{k,t}^{m}$ 表示教师模态相对于学生模态的置信程度，利用超参数 $\gamma_t$ 表示蒸馏在当前通信轮次的启用强度，就得到AMRD蒸馏损失，如式(7)所示：

$$L_{\mathrm{AMRD}}^{k,t}=\min\left(\lambda\kappa\gamma_{t}\cdot\frac{\sum_{m\in S_{k,t}}\omega_{k,t}^{m}L_{\mathrm{rel}}^{k,t,m}}{\sum_{m\in S_{k,t}}\omega_{k,t}^{m}+\varepsilon},\xi\cdot sg\left(L_{\mathrm{cls}}^{k,t}\right)\right) \tag{7}$$

其中，$\lambda$ 控制总体权重，$\kappa$ 调整蒸馏项数值尺度，$\gamma_t$ 为超参数，$\omega_{k,t}^{m}$ 用于表达教师模态的相对可信度. 从优化含义看，$L_{\mathrm{cls}}^{k,t}$ 负责让模型完成当前任务，$L_{\mathrm{AMRD}}^{k,t}$ 负责让弱势模态形成更接近优势模态的样本结构. 二者互补，前者用来约束预测是否正确，后者用来约束模态表征是否稳定. 这种设计使算法可以脱离具体数据模态的限制，不论训练的是图像、文本、音频还是视频数据，只要训练过程中的样本关系在不同模态之间可以进行对齐，就可以构造关系矩阵并进行蒸馏. AMRD算法的单轮流程如图2所示.

**2.5　FedA²适配下的目标函数**　AMRD负责解决客户端本地的跨模态结构补偿问题，在得到本地目标函数后，还要将不同的联邦学习算法对本地训练目标、参数范围和更新方向施加不同约束. 为了避免针对每一种联邦算法单独改写损失函数，本文提出FedA²，将FedAvg，FedProx，FedOpt，FedBN和SCAFFOLD等联邦算法的本地训练过程中的差异统一抽象为本地算法适配项. 不同联邦算法本地训练流程中的差异主要体现在两方面：一是本地目标中是否加入额外约束，二是客户端哪些参数参与后续聚合. 本文将这些差异抽象为算法相关项，形成统一的本地优化形式.

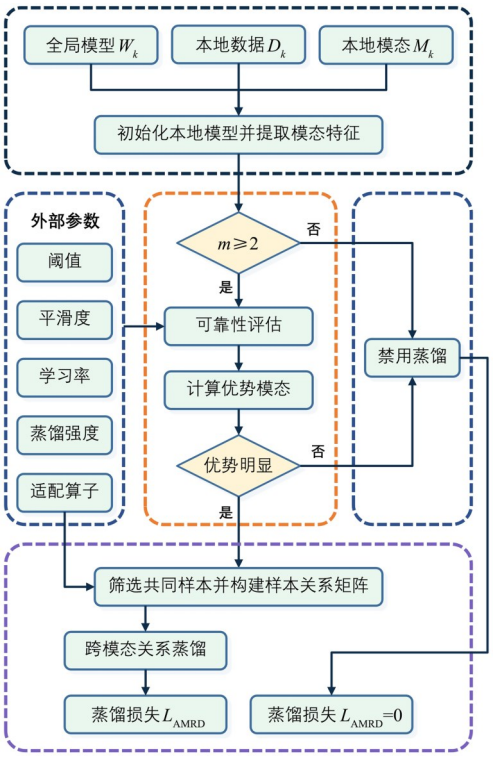



**图 2 单轮 AMRD 算法流程**

**Fig. 2 The single-round AMRD algorithm process**

首先,将传统分类损失和 AMRD 蒸馏损失合并为客户端 $k$ 在第 $t$ 轮的基础增强目标 $G_{k,t}(w_k)$:

$$G_{k,t}(w_k)=L_{\text{cls}}^{k,t}(w_k)+L_{\text{AMRD}}^{k,t}(w_k) \tag{8}$$

式(8)中,分类损失 $L_{\text{cls}}^{k,t}$ 负责保证任务预测能力,蒸馏损失 $L_{\text{AMRD}}^{k,t}$ 负责约束弱模态学习强模态的样本关系结构. 此过程中,FedA²不替代原有分类目标,仅在分类监督之外增加一个模态结构补偿项. 在此基础上,引入统一的算法适配项 $\Omega_A^{k,t}$:

$$\Omega_A^{k,t}(w_k)=\frac{\mu_A}{2}\left\| P_A(w_k-w^t)\right\|_2^2+\left\langle d_A^{k,t},P_A(w_k-w^t)\right\rangle \tag{9}$$

其中,$P_A$ 表示算法参与约束的参数子空间,$\mu_A$ 表示近端约束强度,$d_A^{k,t}$ 表示方向校正项. 前半部分是本地模型与全局模型之间的距离约束,用来刻画 FedProx 类算法对客户端漂移的抑制;后半部分是本地更新方向的修正,用来刻画 SCAFFOLD 类算法对客户端更新方向的校正. 对于 FedAvg,FedOpt 等不需要额外本地正则的算法,可以直接通过参数使对应项自然退化. 通过调节 $P_A$,$\mu_A$ 和 $d_A^{k,t}$,可以把本文研究的 FedAvg,FedProx,FedOpt,FedBN 和 SCAFFOLD 的本地目标统一,如式(10)所示:

$$\min_{w_k} J_A^{k,t}(w_k)=G_{k,t}(w_k)+\Omega_A^{k,t}(w_k) \tag{10}$$

其中,$J_A^{k,t}$ 为经过 FedA² 适配的本地目标. 训练结束后,客户端输出的更新参数 $\tilde{w}_k^{t+1}$ 的统一表示如式(11)所示:

$$\tilde{w}_k^{t+1}=w^t+\zeta_A P_A\left(w_k^{t+1}-w^t\right) \tag{11}$$

在本文研究范围内,无论本地采用哪一种联邦算法,均受 FedA² 中适配项和参数的影响,具体如下.

(1)对于 SCAFFOLD,在引入优化后本地更新方向变为 $\nabla_{w_k}J_A^{k,t}=\nabla_{w_k}G_{k,t}+P_A^T\left(c^t-c_k^t\right)$,其中,$c^t$ 表示全局控制变量,$c_k^t$ 表示客户端 $k$ 的本地控制变量,这与 SCAFFOLD 使用控制变量来修正客户端漂移的思想一致. 对于 FedProx,$\Omega_A^{k,t}$ 退化为正则项.

(2)对于 FedAvg,FedOpt 和 FedBN,$\Omega_A^{k,t}$ 不改变本地损失.

(3)为了进一步兼容 FedOpt 和 FedBN,客户端输出参数为 $\tilde{w}_k^{t+1}=w^t+\zeta_A P_A\left(w_k^{t+1}-w^t\right)$,其中,$\tilde{w}_k^{t+1}$ 表示客户端上传或参与聚合的参数更新结果,$\zeta_A$ 表示算法相关的更新缩放系数. 对于 FedAvg,FedProx 和 SCAFFOLD,通常可取 $\zeta_A=1$;对于 FedOpt,可取 $0<\zeta_A\leqslant 1$ 以控制本地更新幅度;对于 FedBN,$P_A$ 可屏蔽 BN 相关参数,使 BN 参数保留在客户端本地.

这样,在 FedA² 适配的本地目标中,AMRD 可以不依赖某一个具体的联邦优化算法,也不依赖某一个固定不变的模态类型,还可以接受传感器、图像、文本、音频和视频的任意组合形成的多模态数据,很适用于现实的联邦学习场景.

联邦训练过程如以下算法所示.

**算法 自适应多模态蒸馏算法 AMRD**

**输入**:当前轮次全局模型参数 $w^t$,客户端 $k$ 本地数据集 $D_k$,客户端模态集合 $M_k$,联邦算法类型 $A$,训练轮数 $E$ 以及相关超参数

**输出**:客户端 $k$ 在第 $t$ 轮本地训练后得到的模型参数 $\tilde{w}_k^{t+1}$

1. 初始化本地模型:$w_k \leftarrow w^t$
2. for $epoch=1,2,\cdots,E$ do

3. 采样一个本地批次 $B_k \subset D_k$
4. for each $m \in M_k$ do
   提取模态特征 $Z_{k,t}^m = E_m(B_k^m; w_k)$
5. end for
6. 计算多模态预测结果 $\hat{Y}_{k,t} = f_{w_k}(X_{k,t}, M_k)$
7. 由式(2)计算分类损失 $L_{\text{cls}}^{k,t}$
8. 构造候选模态集合 $E_{k,t} = \{m \mid m \in M_k\}$
9. if $|E_{k,t}| < 2$ 判断是否启用AMRD
   $L_{\text{AMRD}}^{k,t} = 0$
10. else
11. for each $m \in E_{k,t}$ do
    根据式(3)计算可靠性分数 $s_{k,t}^m$
    更新平滑分数 $\bar{s}_{k,t}^m$
12. end for
13. 选择教师模态 $m_{k,t}^+ = \text{argmax}_{m \in E_{k,t}} \bar{s}_{k,t}^m$
14. 构造学生模态集合
    $S_{k,t} = \left\{ m \middle| m \in E_{k,t}, m \neq m_{k,t}^+, \bar{s}_{k,t}^{m_{k,t}^+} - \bar{s}_{k,t}^m \geqslant \delta \right\}$
15. for each $m \in S_{k,t}$ do
    构造共同样本集合 $J_{k,t}^m$
    由式(5)计算 $\boldsymbol{R}_{k,t}^m$ 和 $\boldsymbol{R}_{k,t}^+$
    由式(6)计算 $L_{\text{rel}}^{k,t,m}$
16. end for
17. 由式(7)计算AMRD蒸馏损失 $L_{\text{AMRD}}^{k,t}$
18. end if
19. 适配算子 $(P_A, \mu_A, d_A^{k,t}, \zeta_A) \leftarrow A(w^t, w_k, c^t, c_k^t)$
20. 构造算法正则项
    $$\Omega_A^{k,t}(w_k) = \frac{\mu_A}{2} \| P_A(w_k - w^t) \|_2^2 + \left\langle d_A^{k,t}, P_A(w_k - w^t) \right\rangle$$
21. 构造AMRD增强后的统一本地目标
    $J_A^{k,t}(w_k) = L_{\text{cls}}^{k,t}(w_k) + L_{\text{AMRD}}^{k,t}(w_k) + \Omega_A^{k,t}(w_k)$
22. 更新本地模型 $w_k \leftarrow w_k - \eta \nabla_{w_k} J_A^{k,t}(w_k)$
23. 生成聚合参数 $\tilde{w}_k^{t+1} = w^t + \zeta_A P_A(w_k^{t+1} - w^t)$
24. end for
25. return $\tilde{w}_k^{t+1}$

## 3　实验设置与分析方案

**3.1　数据集与模态类型**　为了检验AMRD与FedA²在不同模态组合和不同任务场景下的适应性，采用UCI-HAR，MHealth，HHAR，CREMA-D和Crisis-MMD五个数据集进行实验，覆盖视频、图像、文本、传感器和音频等典型模态类型，同时包含真实采集环境中常见的主体差异和设备差异，更贴近多模态联邦学习的应用场景.

**3.2　联邦算法设置**　选择FedAvg，FedProx，FedOpt，FedBN和SCAFFOLD作为对比算法，覆盖联邦优化的几种典型机制，分别从本地目标约束、更新方向修正、参数作用范围和更新使用方式影响联邦训练过程.

FedProx近端正则项设置为0.001；FedOpt服务端采用Adam优化器，服务端的初始学习率设置为0.025；SCAFFOLD中全局控制变量和客户端控制变量初始化为0，服务端初始学习率设置为1.0，本地学习率与FedAvg保持一致.

每种联邦算法均设置基础版本和接入AMRD-FedA²(A-F)的版本，通过对比实验来测试AMRD的性能，同时分析FedA²的鲁棒性.

**3.3　扰动设置**　借助FedMultimodal多模态联邦实验项目[22]，可以手动设置模态缺失、标签缺失和错误标签三类扰动，来模拟现实世界中数据不完整和监督不足的问题. 其中，模态缺失模拟部分输入模态不可用的情况，标签缺失模拟监督信号不足的情况，错误标签模拟标注噪声或类别误标现象.

与模态缺失和标签缺失不同，错误标签不是信息缺失，而是信息污染. 错误标签会同时影响分类损失和模态可靠性评分，会导致教师模态选择错误，而本文提出的方法会进一步放大错误选择带来的负提升. 正确解决错误标签问题的原理更接近噪声标签鲁棒学习，需要结合标签纠错、样本重加权和鲁棒损失函数等额外机制，这和本文研究方案属于两个不同的领域. 因此，本文的实验扰动围绕模态缺失和标签缺失，不对错误标签扰动进行深究.

扰动强度采用0.1～0.5的等差数列，0.4及以上可以表示为重度扰动场景，用于检验方法在极端模态缺失或标签缺失场景下的稳定性. 数据异质性划分主要采用Dirichlet分布参数($\alpha = 0.1$)来构造高度non-IID场景，实验中采用($\alpha = 5.0$)作为弱异质性对照.

通过不同扰动强度的对比，可以分析AMRD是否存在对教师模态过度依赖、蒸馏信号不稳定

或弱模态补偿不足等问题,证明算法在现实场景中的鲁棒性.

**3.4 实验环境和性能指标** 每个实验测试用例均在NVIDIA RTX 4090 GPU上训练至收敛.主要性能指标采用$F1$分数和$Acc$(Accuracy),初始学习率设置为0.005,每个不同的测试项都经过200个epoch.实验结果如表2和表3所示,表中黑体字表示性能更优,$\alpha$为Dirichlet分布参数,$mm$为模态缺失率,$ml$为标签缺失率,Clear为未开启AMRD-FedA$^2$,A-F为同条件下开启AMRD-FedA$^2$.

## 4 实验结果与分析

**4.1 不同联邦算法下的整体性能** 观察引入A-F对实验结果的影响.在高度非独立同分布条件下,统计五种代表性联邦学习算法在五个多模态数据集上的性能变化.在模态缺失率50%和标签缺失率50%下,分别给出Clear基线和启用A-F后的$F1$.结果显示,在高度非独立同分布和高度缺失率下,A-F几乎在所有组合上均取得正向提升,具有较好的跨算法和跨模态任务泛化能力.

由表2可见,在UCI-HAR数据集上,各算法的$F1$均有提升.其中,FedOpt从0.6027提升至0.7576,增幅最大;在MHealth数据集上,FedAvg和FedProx分别提升0.1907和0.1833;在HHAR数据集上,SCAFFOLD从0.6072提升至0.7401;而在文本图像模态任务Crisis-MMD数据集上,算法的提升最显著.说明当图文模态互补性较强且缺失模态造成较大语义损失时,所提方法能够显著改善模型性能.

由表3可见,大多数算法均获得了明显提升.MHealth数据集更是在FedAvg,FedProx,FedBN和SCAFFOLD上分别提升了0.2089,0.2629,0.2068和0.3722,提升较大.表明在模态信息仍然完整但标签稀缺的情况下,本文方法能够缓解因标签不足导致的性能退化问题.

需要注意FedOpt部分退化和SCAFFOLD基线性能较低的情况.FedOpt在联邦学习研究中属于较强基线算法,在客户端更新方向稳定时服务端的优化器能够提升收敛效率.但在高缺失场景下,客户端更新包含较强噪声和结构性偏差,此时A-F引入的局部模态适配信号虽然能够改善部分客户端表示,但也可能改变本地更新方向,使FedOpt服务端统计受到干扰,增强机制未必持续

**表2 $\alpha$=0.1及$mm$=0.5时各数据集在不同联邦算法上的$F1$**

**Table 2 Results of each dataset on different federated algorithms when $\alpha$=0.1 and $mm$=0.5**

| 方法 | UCI-HAR | | CREMA-D | | MHealth | | HHAR | | Crisis-MMD | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Clear | A-F | Clear | A-F | Clear | A-F | Clear | A-F | Clear | A-F |
| FedAvg | 0.6803 | 0.7223 | 0.5537 | 0.5731 | 0.7250 | **0.9157** | 0.7274 | 0.7923 | 0.1742 | **0.3231** |
| FedProx | 0.6918 | 0.7151 | 0.5618 | 0.5949 | 0.7581 | **0.9414** | 0.7735 | 0.8036 | 0.2041 | **0.3515** |
| FedOpt | 0.6027 | **0.7576** | 0.5749 | 0.5920 | 0.7153 | 0.8103 | 0.6921 | **0.7659** | 0.1189 | **0.3046** |
| FedBN | 0.6971 | 0.7270 | 0.5847 | 0.5911 | 0.7389 | 0.8392 | 0.7104 | 0.7483 | 0.1723 | **0.3829** |
| SCAFFOLD | 0.4489 | **0.5283** | 0.5012 | **0.5745** | 0.6926 | 0.8361 | 0.6072 | **0.7401** | 0.0845 | **0.3489** |

**表3 $\alpha$=0.1及$ml$=0.5时各数据集在不同联邦算法上的$F1$**

**Table 3 Results of each dataset on different federated algorithms when $\alpha$=0.1 and $ml$=0.5**

| 方法 | UCI-HAR | | CREMA-D | | MHealth | | HHAR | | Crisis-MMD | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Clear | A-F | Clear | A-F | Clear | A-F | Clear | A-F | Clear | A-F |
| FedAvg | 0.4396 | **0.5879** | 0.5707 | 0.5913 | 0.6930 | 0.9019 | 0.7401 | 0.7720 | 0.1136 | 0.3530 |
| FedProx | 0.4580 | **0.5704** | 0.5716 | 0.5960 | 0.6716 | 0.9345 | 0.7682 | 0.8462 | 0.1239 | 0.3419 |
| FedOpt | 0.7604 | 0.7212 | 0.5962 | 0.5893 | 0.8303 | 0.9201 | 0.8129 | 0.8501 | 0.1989 | 0.3870 |
| FedBN | 0.4101 | 0.5302 | 0.5677 | 0.5801 | 0.7047 | 0.9115 | 0.7547 | 0.7709 | 0.1023 | 0.3225 |
| SCAFFOLD | 0.2763 | **0.4795** | 0.5014 | **0.5721** | 0.5031 | **0.8753** | 0.6370 | **0.8105** | 0.0845 | **0.3313** |

带来收益，甚至可能造成轻微退化. SCAFFOLD依赖本地梯度估计的可靠性，在高度模态缺失或标签缺失场景下，客户端梯度极易受影响，导致Clear基线较低. A-F可在一定程度上改善客户端更新质量，因此SCAFFOLD在多组实验中都有较大的相对提升.

**4.2　消融实验分析**　实验中A-F算法模块部分设计了独立的启用开关，在保持主网络结构、训练策略及其他实验参数完全一致时，通过改变A-F模块的启用状态来观察算法对模型性能的贡献.

由表2和表3还可知，在启用A-F算法模块后，不同联邦优化器的模型精度指标均发生了明显变化. 除了个别数据因优化器的特性而出现异常外，其余指标均表现出不同程度的提升，且提升趋势具有较好的一致性. 表明A-F模块能够在联邦训练过程中持续发挥正向优化作用，验证了算法在整体联邦训练过程中的实际贡献.

**4.3　扰动场景分析**

**4.3.1　不同模态缺失率和标签缺失率的场景**　为了进一步分析算法在不同扰动强度下的鲁棒性，对比模态缺失率和标签缺失率从0增加到50%时，在各扰动场景下Clear基线的性能有提升. 以UCI-HAR数据集为例，其结果如图3和图4所示. 由图可见，在模态缺失场景与标签缺失场景中，随着缺失率逐渐升高，A-F的提升幅度整体呈扩大趋势. 当缺失率较低时，各算法的提升相对有限，缺失率超过30%后，多数算法的增益明显增加. 说明在模态信息和标签信息受损更加严重时，算法更能发挥弱模态补偿作用，可显著缓解模态缺失或标签缺失导致的性能退化. 因此，本文方法更适合高模态缺失和高标签缺失的真实多模态联邦学习困难场景.

此外，FedOpt在模态缺失场景中提升较大，但在标签缺失场景中存在负向波动. 这种差异来自FedOpt本身特性. 高标签缺失场景下，客户端本地梯度估计方差增大，部分类别可能在本地训练中长期缺失，导致更新方向极不稳定，A-F的额外局部更新会进一步加剧这种现象，最终表现为负增长. 但这没有否定A-F方法的整体有效性，只是反映不同联邦优化器对高标签缺失扰动的适应能力存在差异.

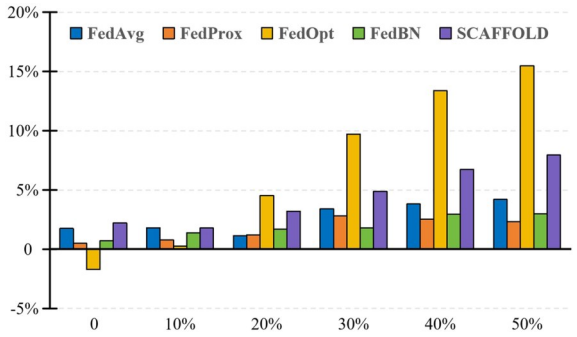



**图3　不同模态缺失率下和Clear相比的*Acc*提升**

**Fig. 3　*Acc* improvement compared to Clear under different modality missing rates**

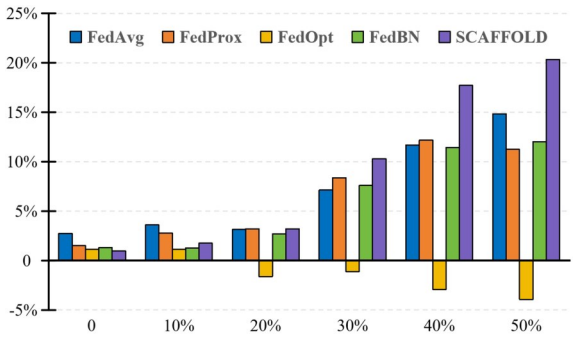



**图4　不同标签缺失率下和Clear相比的*Acc*提升**

**Fig. 4　*Acc* improvement compared to Clear under different label missing rates**

**4.3.2　不同non-IID程度的场景**　以Crisis-MMD文本-图像双模态数据集为例，对比$\alpha=0.1$和$\alpha=5.0$两种数据划分条件下，各联邦算法在启用A-F前后的性能变化. 其中，$\alpha=0.1$表示客户端间数据分布差异较大，属于高度non-IID场景；$\alpha=5.0$表示客户端数据分布更接近IID，异质性相对较弱. 实验结果如图5所示. 由图可见，$\alpha=0.1$时，A-F对各联邦算法均带来明显提升，*Acc*

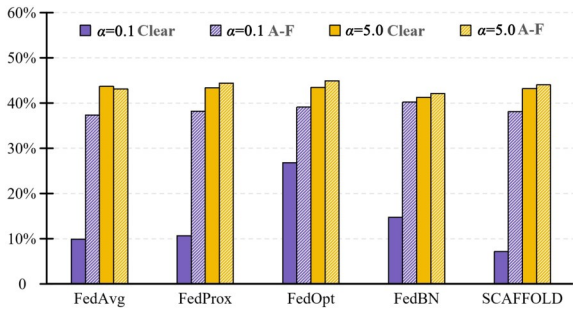



**图5　不同non-IID程度下的*Acc*对比**

**Fig. 5　*Acc* with different non-IID levels**

显著高于Clear基线. 在$\alpha=5.0$的近IID场景下,由于各算法的Clear基线本身已经较高,故启用A-F后整体性能与基线接近. 说明当客户端之间存在较强数据分布差异时,单纯依赖原始联邦聚合容易受到局部数据偏移的影响,而A-F能够通过自适应模态关系建模和联邦算法适配来缓解异质数据导致的全局模型退化. 当客户端数据分布较一致时,原始联邦算法已经能够较好地学习全局模型,额外自适应增强机制带来的边际收益会减小.

综上,客户端原始数据的异质性越强,本文提出的方法对训练效果的影响程度越大,尤其适合客户端本地存在高度异质的数据场景.

## 5 结论

本文针对多模态联邦学习中的数据异质、模态缺失与标签缺失问题,提出AMRD-FedA$^2$框架,通过自适应模态关系蒸馏与联邦算法适配机制,提升模型在复杂扰动场景下的鲁棒性,具有一定的泛化能力.

实验证明,该方法在不同模态、不同联邦算法以及non-IID条件下均具有稳定的性能,在高度数据异质场景下拥有较好的表现.

未来可进一步研究更细致的模态可靠性评估方案和低开销的跨客户端知识协同与隐私保护机制,提升算法在真实场景下的应用前景.

### 参考文献


[1] Kairouz P, McMahan H B, Avent B, et al. Advances and open problems in federated learning. Foundations and Trends in Machine Learning, 2021, 14(1/2): 1－210.

[2] 高莹,陈晓峰,张一余,等. 联邦学习系统攻击与防御技术研究综述. 计算机学报,2023,46(9):1781－1805.

[3] Che L W, Wang J Q, Zhou Y, et al. Multimodal federated learning: A survey. Sensors, 2023, 23(15): 6986.

[4] 林伟伟,石方,曾岚,等. 联邦学习开源框架综述. 计算机研究与发展,2023,60(7):1551－1580.

[5] Le H Q, Nguyen M N H, Thwal C M, et al. FedMEKT: Distillation-based embedding knowledge transfer for multimodal federated learning. Neural Networks, 2025, 183: 107017.

[6] Wang S, Qu Z, Liu Y, et al. FedMMR: Multi-modal federated learning via missing modality reconstruction// 2024 IEEE International Conference on Multimedia and Expo. Niagara Falls, Canada: IEEE, 2024: 1－6.

[7] Chen J Y, Zhang A D. FedMSplit: Correlation-adaptive federated multi-task learning across multimodal split networks// Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. New York, NY, USA: Association for Computing Machinery, 2022: 87－96.

[8] Ouyang X M, Xie Z Y, Fu H M, et al. Harmony: Heterogeneous multi-modal federated learning through disentangled model training// Proceedings of the 21st Annual International Conference on Mobile Systems, Applications and Services. New York, NY, USA: Association for Computing Machinery, 2023: 530－543.

[9] Yu Q Y, Liu Y, Wang Y M, et al. Multimodal federated learning via contrastive representation ensemble// The 11th International Conference on Learning Representations. Kigali: ICLR, 2023, arXiv: 2302.08888.

[10] Le H Q, Thwal C M, Qiao Y, et al. Cross-modal prototype based multimodal federated learning under severely missing modality. Information Fusion, 2025, 122: 103219.

[11] 周弈志,王军晓,谢鑫,等. 联邦长尾学习研究综述. 计算机学报,2025,48(4):779－807.

[12] Gou J P, Yu B S, Maybank S J, et al. Knowledge distillation: A survey. International Journal of Computer Vision, 2021, 129(6): 1789－1819.

[13] 潘海为,于丰铭,张可佳,等. 面向视觉算法的知识蒸馏研究综述. 计算机研究与发展,2026,63(1):90－122.

[14] Park W, Kim D, Lu Y, et al. Relational knowledge distillation// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA: CVF, 2019: 3967－3976.

[15] Tung F, Mori G. Similarity-preserving knowledge distillation// 2019 IEEE/CVF International

Conference on Computer Vision. Seoul, Korea (South):IEEE, 2019:1365－1374.

[16] Guo J L, Fu Y J, Zhai Z W, et al. MASA: Multimodal federated learning through modality-aware and secure aggregation. IEEE Transactions on Mobile Computing,2025,24(8):7328－7344.

[17] Mcmahan B, Moore E, Ramage D, et al. Communication-efficient learning of deep networks from decentralized data//Proceedings of the 20th International Conference on Artificial Intelligence and Statistics. New York,NY,USA:PMLR,2017:1273－1282.

[18] Li T, Sahu A K, Zaheer M, et al. Federated optimization in heterogeneous networks// Proceedings of Machine Learning and Systems. Austin,TX,USA:MLSys,2020:429－450.

[19] Karimireddy S P, Kale S, Mohri M, et al. SCAFFOLD: Stochastic controlled averaging for federated learning//Proceedings of the 37th International Conference on Machine Learning. New York,NY,USA:PMLR,2020:5132－5143.

[20] Li X X, Jiang M R, Zhang X F, et al. FedBN: Federated learning on non-IID features via local batch normalization//International Conference on Learning Representations. Vienna, Austria: ICLR, 2021:1－27.

[21] Reddi S J, Charles Z, Zaheer M, et al. Adaptive federated optimization//International Conference on Learning Representations. Vienna, Austria: ICLR, 2021:1－38.

[22] Feng T T, Bose D, Zhang T, et al. FedMultimodal: A benchmark for multimodal federated learning// Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. New York, NY, USA: Association for Computing Machinery,2023:4035－4045.

(责任编辑　杨可盛)