Beijing Forest Studio
北京理工大学信息系统及安全对抗实验中心

面向数据异构与通信高效的联邦大模型优化与应用研究

博士研究生 陈星星

2025年10月26日

问题回溯

- 相关内容
  - 2024.12.19 贺晨阳：《大语言模型的越狱攻击》
  - 2024.11.27 刘栋涵：《基于大语言模型的事件根因分析》
  - 2024.01.03 徐程柯：《大语言模型调研》
  - 2023.04.09 杨得山：《联邦学习的后门防御方法》
  - 2022.08.30 杨得山：《联邦学习的后门攻击方法》
  - 2022.05.16 郝靖伟：《联邦学习及其后门攻击方法初探》

内容提要

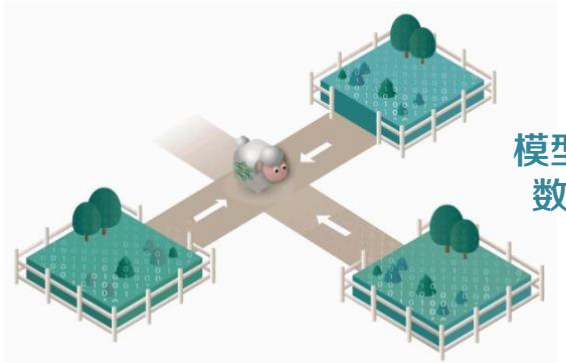
- 预期收获
- 题目内涵解析
- 研究背景与意义
- 研究历史与现状
- 知识基础
- 算法原理
  - Data Divergence-aware Client Selection via Knowledge Graph for Federated LLM Fine-tuning
  - LLM-driven Medical Report Generation via Communication-efficient Heterogeneous Federated Learning
- 特点总结与工作展望
- 参考文献

预期收获

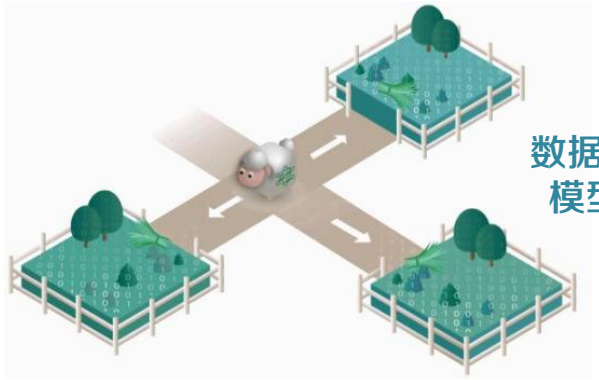
- 预期收获
  - 了解联邦大模型基本原理
  - 理解联邦大模型参数高效微调过程
  - 掌握个性化联邦大模型训练方法

目标内涵—面向数据异构与通信高效的联邦大模型优化与应用研究

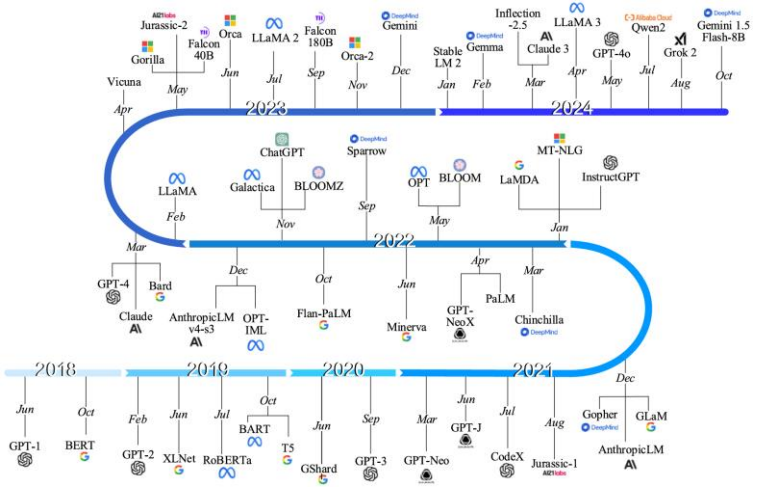
- 研究目标
  - 建立数据异构感知的联邦大模型优化机制
  - 设计差异感知的客户端选择与无偏聚合策略
  - 构建通信高效的联邦大语言模型微调框架
  - 验证在高异构医疗场景下的应用有效性
- 内涵解析
  - 联邦学习：允许各客户端在不共享本地数据的情况下协作训练全局模型
  - 大语言模型：一类基于深度学习的人工智能模型，具备强大的自然语言理解和生成能力



模型不动
数据动



数据不动
模型动



研究历史

2017

McMahan等人提出联邦学习概念与FedAvg算法，明确了在多客户端本地训练与服务端聚合的框架、通信-本地迭代的设计范式

2018

FedProx算法通过引入近端项来保对这些未完成计算的信息进行聚合，从而减少因为设备异构和数据异构问题造成的偏差

2020

Scaffold算法证明了客户端数据中的异质性会导致本地更新中的“漂移”，并结合控制变量和控制梯度的概念，减轻由数据异构引起的模型训练的不稳定性

2022

Wang等人提出一个FedKC模块来解决多语言联邦自然语言理解中的数据异质性挑战，在不共享原始数据的情况下在客户端之间交换知识

2023

FATE-LLM提出一个用于大型语言模型的工业级联合学习框架，集成PEFT、隐私机制与企业场景的工程化支持，提升训练效率的同时保护数据隐私

2024

FeDeRA等算法提出面向联邦LLM的高效微调、权重分解或改进聚合方案，以缓解“桶效应”、异构客户端能力差异等问题

2024

Zeng等人针对医学图像分析中的域偏移问题，设计了一种梯度匹配联合域自适应方法GM-FedDA，以提升FL在跨机构脑图像分类任务中的泛化能力

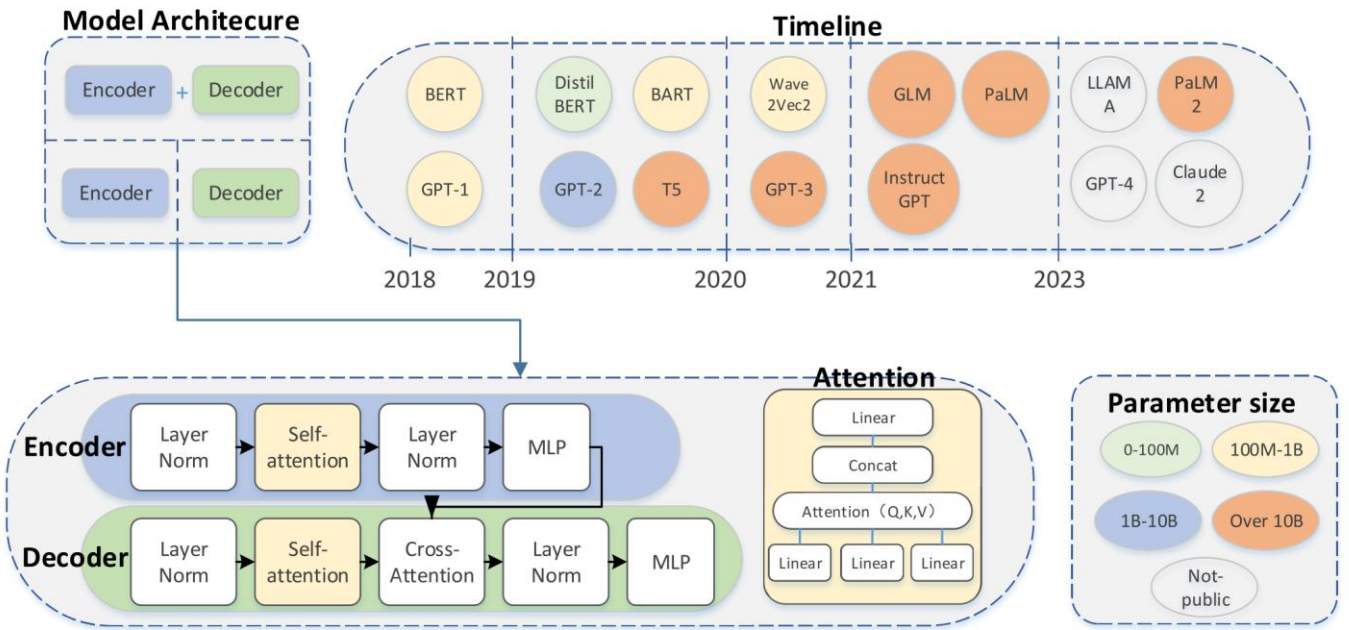
2025

FedALoRA算法提出高效参数微调与个性化聚合算法结合，在解决数据异构性的同时极大的降低了计算开销

2025

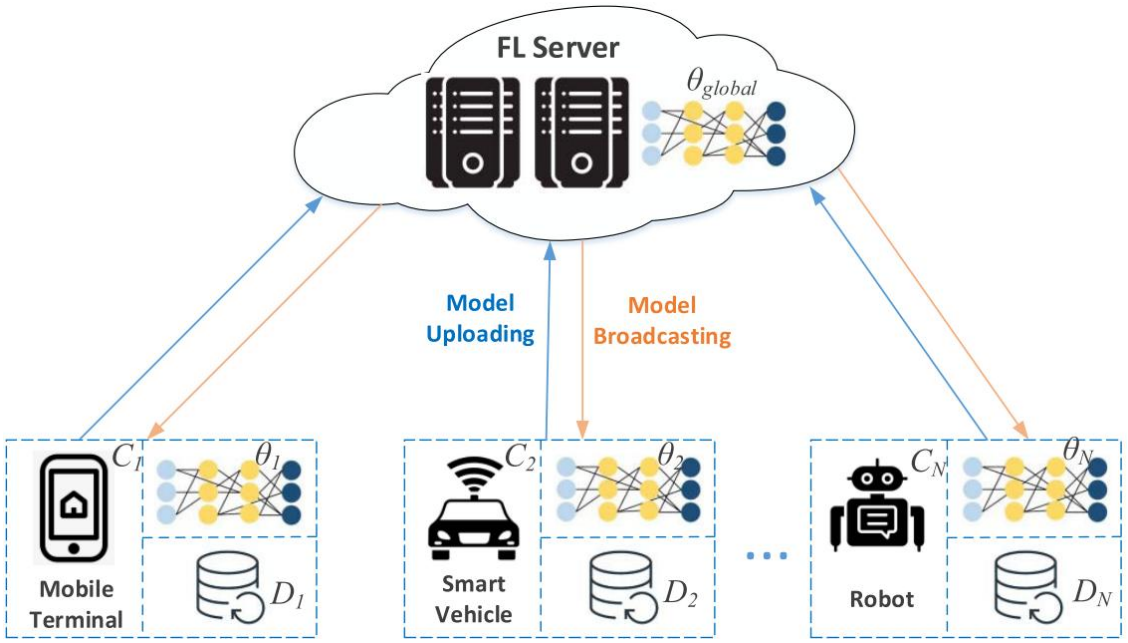
FedSRD算法通过对LoRA参数进行稀疏化-重构-分解处理，实现了通信成本降低90%的突破

- 大语言模型（Large Language Model）
  - 在大规模语料上训练、包含百亿级别（或更多）参数的语言模型，例如GPT，DeepSeek，LLaMA等
  - 目前的大语言模型采用与小模型类似的Transformer架构和预训练目标，与小模型的主要区别在于增加模型大小、训练数据和计算资源
  - 涌现能力，代表性的涌现能力包括上下文学习、指令遵循、逐步推理等

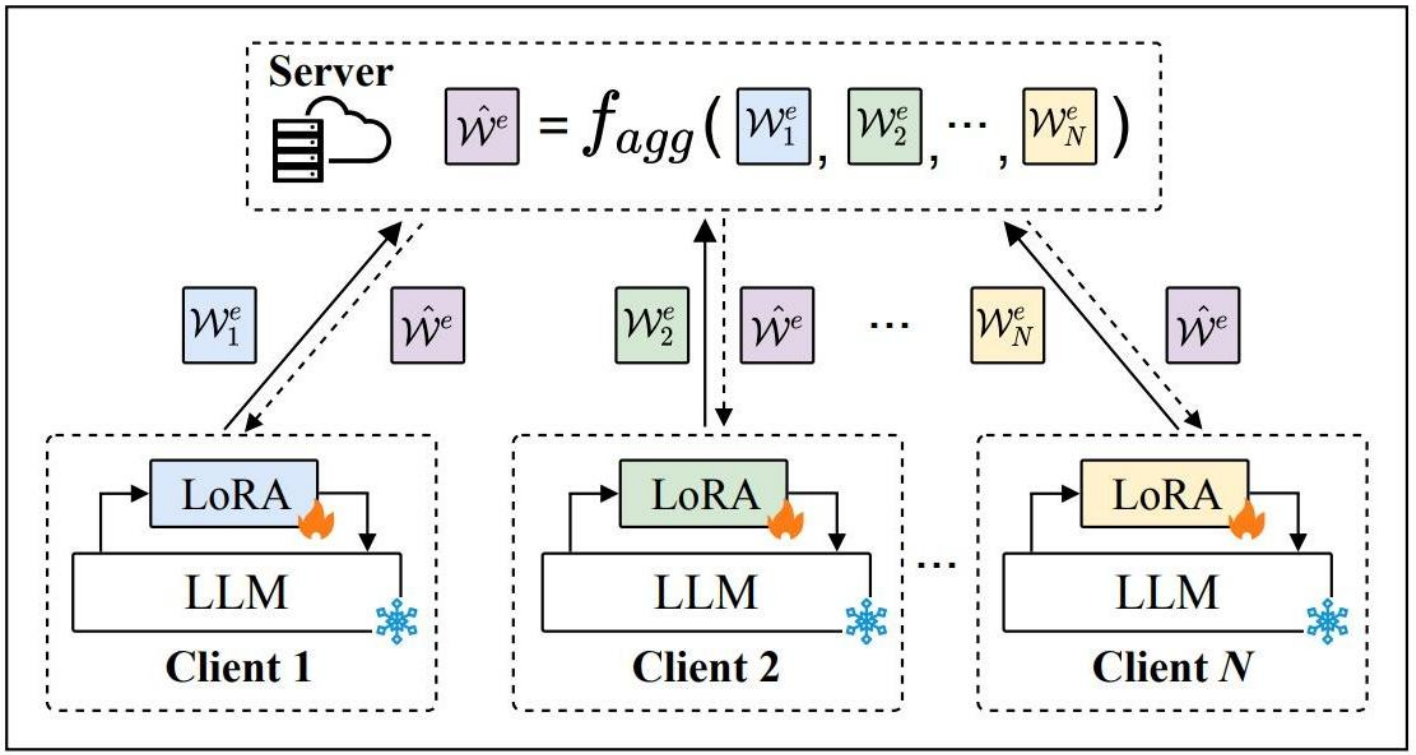


- 联邦学习（Federated Learning）
  - 一种分布式机器学习框架，在不集中原始数据的前提下，通过各参与节点（客户端）的协同训练，构建一个共享的全局模型（服务器）
  - 典型聚合方式为：

$$W^t = \sum_{i=1}^{N} \frac{|D_i|}{|D|} W_i^t$$



- 联邦大语言模型
  - 在数据分散、互不共享的前提下，协同训练强大的大语言模型
  - 模式：数据不动模型动，只交换模型参数，不移动原始数据



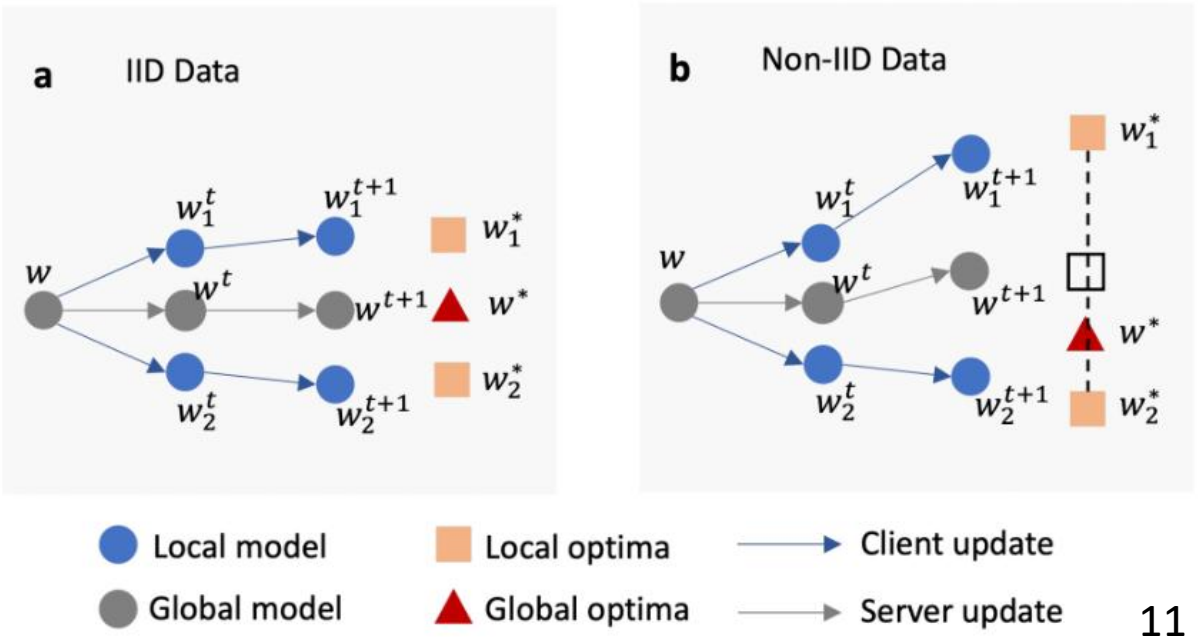
- 联邦模型VS联邦大语言模型

对比维度	联邦模型	联邦大模型
模型类型	通常为传统的轻量级模型	参数量巨大的大语言模型
核心目标	保护数据隐私，解决“数据孤岛”问题，完成特定任务（如图像分类、推荐）	在保护隐私的前提下，训练或微调基础性、通用性的强大语言模型。
通信与计算开销	相对较低	极高
关键技术重点	传统的联邦平均算法 应对非独立同分布数据 基础的差分隐私或安全聚合	参数高效微调，降低开销 更复杂的异构性处理和个性化技术 对安全聚合算法的效率和规模要求更高
数据异构性处理	主要关注样本或特征分布的差异	面临更复杂的概念/语义层面的异构
典型应用场景	手机输入法预测 医疗影像联合分析 跨银行联合风控	多中心联合训练的生物医学模型 金融机构联合微调行业专属模型 利用边缘设备数据改进个人助理

知识基础　什么是异质性

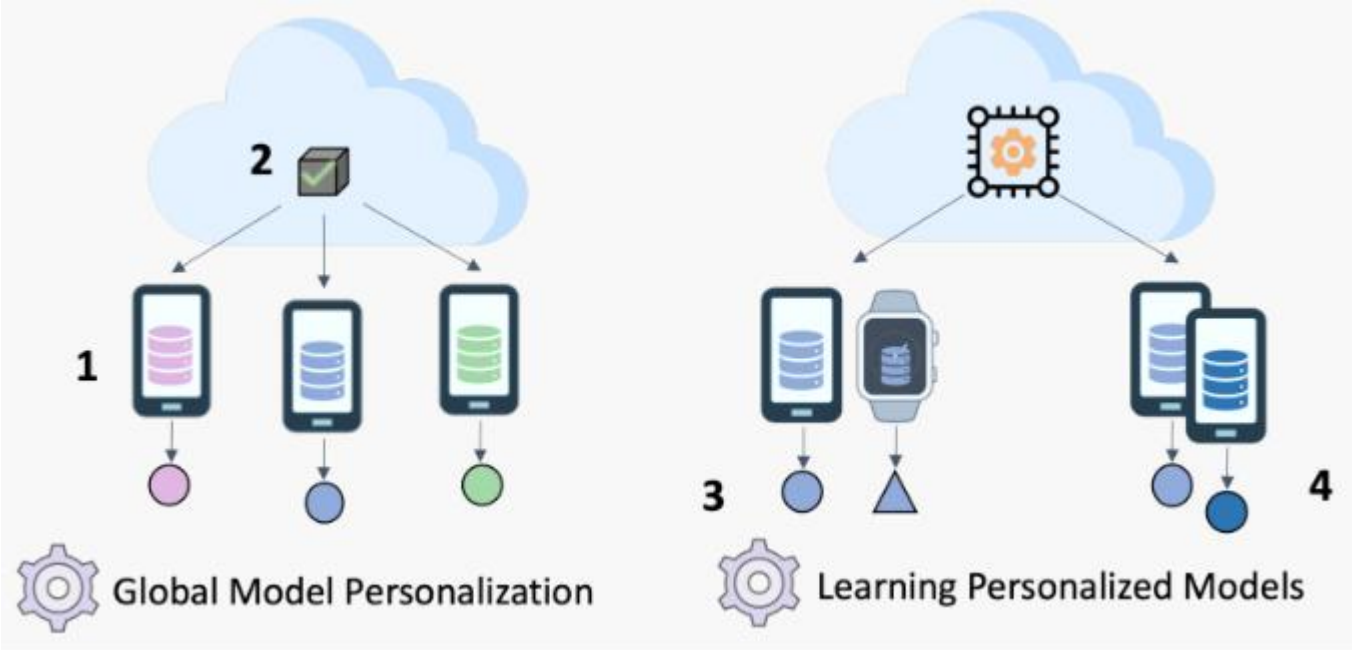
- 独立同分布（IID）
  - 独立：每一个随机变量的取值不受其他随机变量取值的影响
  - 同分布：随机变量遵从同样的概率分布
- 非独立同分布（Non-IID）
  - 一组随机变量不满足独立或同分布的条件，或两者都不满足（时间序列数据）

- 数据异质性
  - 不同客户端收集数据类型不同，导致数据分布不一致
- 数据分布不均衡
  - 某些客户端可能有大量数据，而其他客户端的数据较少，数据标签分布也有所不同

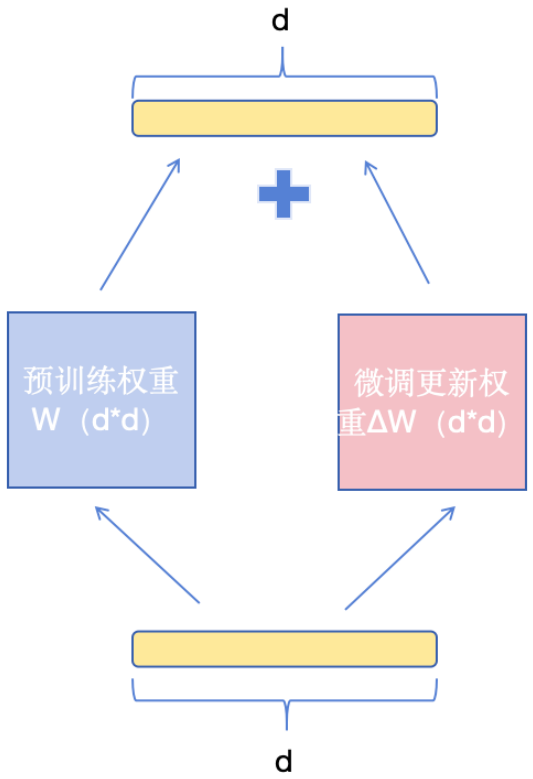


知识基础　个性化联邦学习

- 个性化联邦学习（PFL）
  - 提出在共享全局知识的同时，为每个客户端学习定制化模型，实现“协同中保留差异”
    - 全局模型个性化，提升在异质数据上联邦训练的全局共享模型的性能
    - 学习个性化的模型，提供个性化解决方案



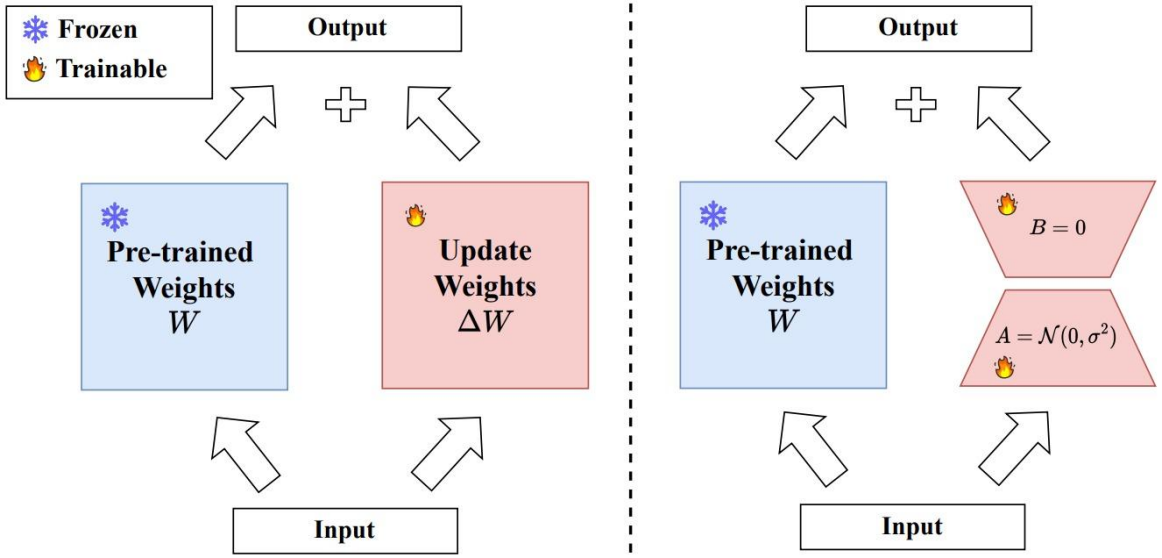
- 参数高效微调（PEFT）
  - 冻结住大模型的主干参数，引入一小部分可训练的参数作为适配模块进行训练
  - 通过微调少量（额外）模型参数或者减少迭代次数，可以使LLM适应下游任务，在不影响任务性能的情况下大幅降低计算需求，节省模型微调时的显存和参数存储开销，降低微调成本



- 低秩适应（LoRA）

$$h = Wx \rightarrow Wx + \Delta Wx = Wx + W_B W_A x$$

其中$\Delta W \in R^{d\times k}$，$W_B \in R^{d\times r}$，$W_A \in R^{r\times k}$，$r \ll \min(d,k)$。在LLM权重矩阵$W$上引入低秩适配器$\Delta W = W_B W_A$，训练时只更新低秩矩阵，从而大幅减少需要上传/下载的参数量（通信量与计算量下降）



Data Divergence-aware Client Selection via Knowledge Graph for Federated LLM Fine-tuning

T	目标	在联邦大语言模型微调场景中，同时优化系统异质性与统计异质性，以减少总训练延迟并提升模型精度
I	输入	各客户端本地语料数据集：MedQA(包含 61,097 题)、MedMCQA(194,000道题目)、MMLU-Medical(15908道题目)、MIMIC-III(53,423例数据)；客户端数据生成的知识图谱、LLama2-7b、Mistral-7b预训练大语言模型参数
P	处理	1. 联邦优化建模 2. 知识图谱辅助异质性度量 3. 优化求解与聚合策略
O	输出	每个客户端被选中的概率、聚合后全局LLM微调参数、性能指标

P	问题	1.系统异质性导致通信瓶颈，统计异质性导致模型收敛不稳定 2. LLM参数量大，在线梯度测量不现实
C	条件	同步联邦学习场景（需等待最慢客户端）、每个客户端拥有局部文本数据，可独立生成知识图谱
D	难点	1. 不同客户端语料语义差异复杂，需设计结构化度量方法 2.如何在离线计算语义差异的同时动态调整客户端采样概率
L	水平	2025 中科院1区Top

知识基础 知识图谱（KG）

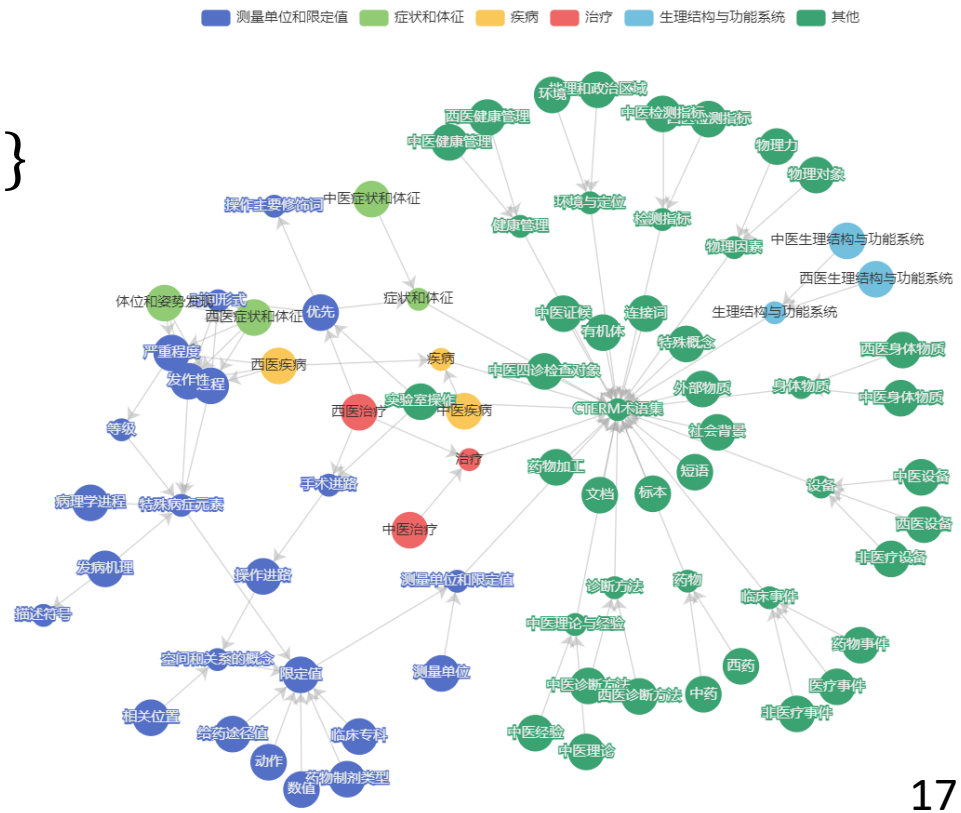
- 知识图谱
  - 一种用于表示和组织现实世界知识的语义网络结构
  - 用节点表示实体，用边表示实体间的关系，并通过属性与语义规则实现对知识的结构化与可计算化表达
  - 将文本抽象为实体-关系-实体三元组

$$G = \{(h, r, t) | h, t \in \varepsilon, r \in R\}$$

- 知识图谱辅助的异质性建模
  - 文本 → 三元组抽取 (h,r,t)
  - KG相似度计算：

$$div(i,j) = 1 - \frac{|KG_i \cap KG_j|}{|KG_i \cup KG_j|}$$

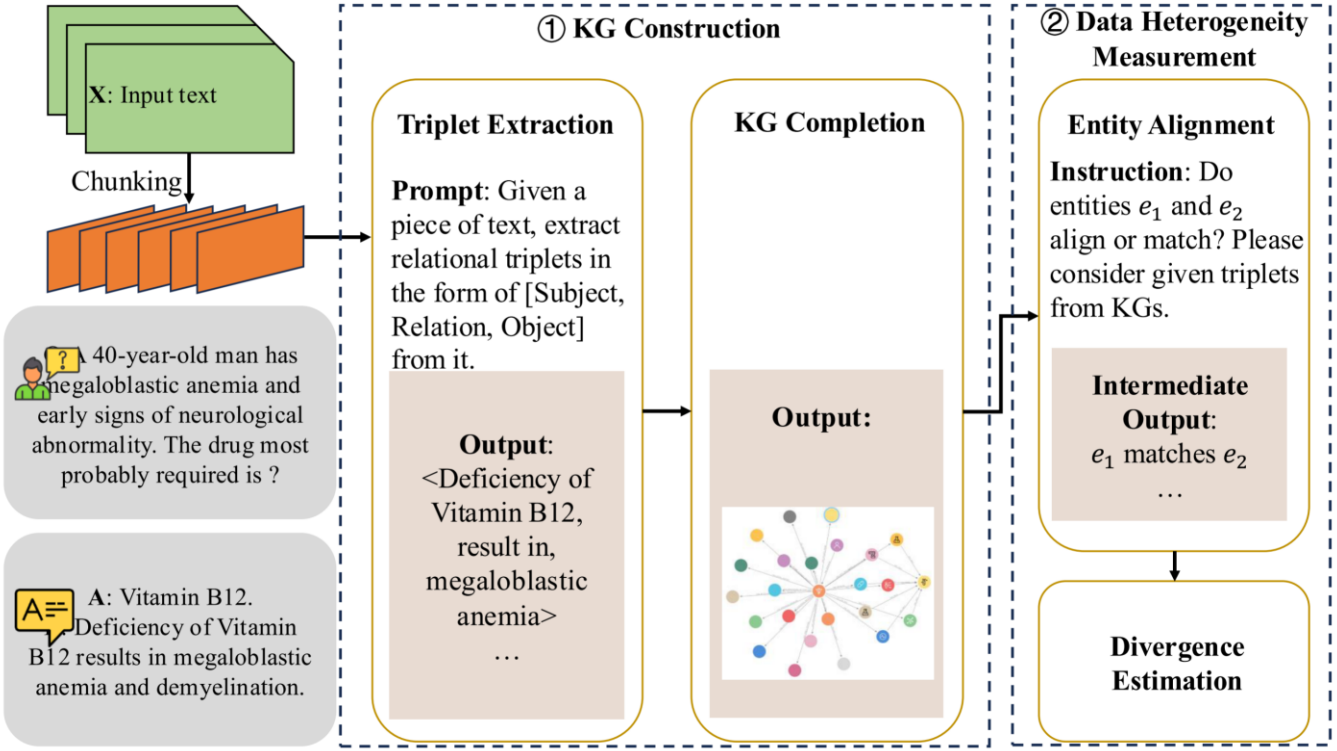
  - 构建数据异质性矩阵
  - 用于采样概率约束



DDCS 算法原理

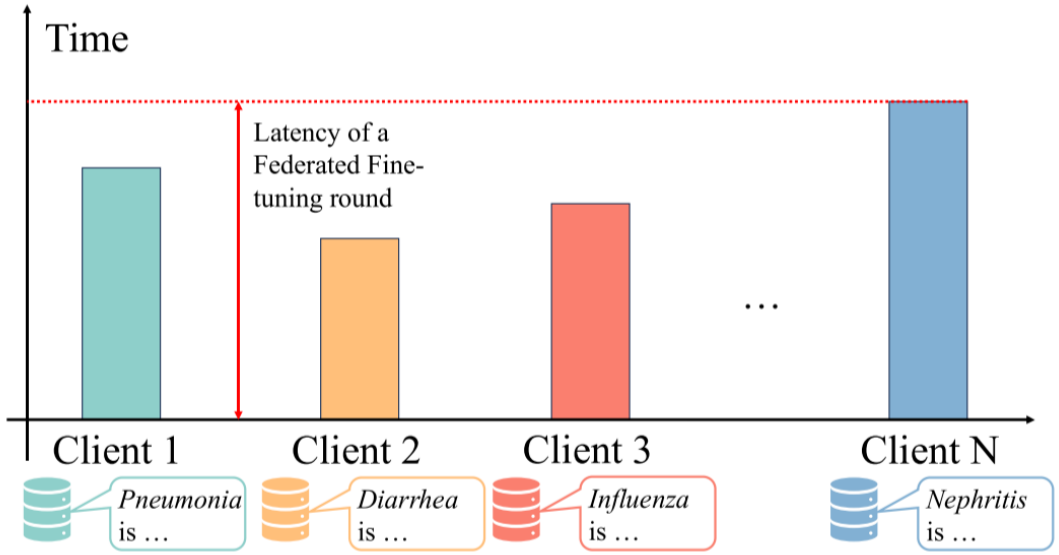
- 算法原理图
  - 离线阶段计算各客户端数据的KG表示并量化“数据散度”
  - 在线训练时依据系统延迟与数据异质性联合优化客户端采样概率

- 算法流程
  - 通过公共节点将三元组连接在一起形成原始知识图
  - 进行知识图补全，预测节点之间缺失的连接，保证KG的完备性
  - 对构造的KG进行对齐并估计数据散度



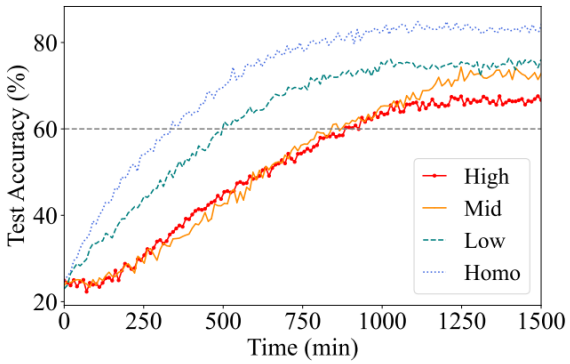
DDCS 创新点分析

- 现有方法存在问题
  - 无效的客户端选择导致：
    - 训练延迟增加（受慢客户端影响）
- 解决方法
  - 首次将知识图谱（KG）引入LLM联邦学习
    - 知识图谱作为一个结构化的语义网络，将每个客户端的数据映射到实体和关系上

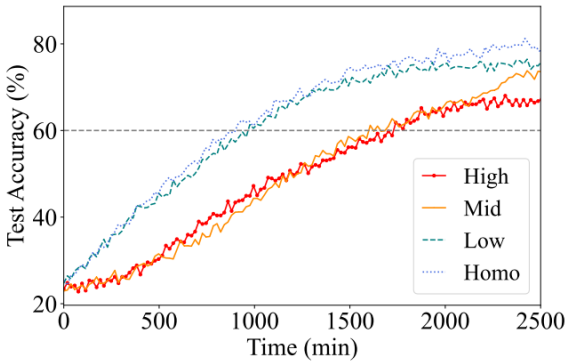


DDCS 创新点分析

- 现有方法存在问题
  - 无效的客户端选择导致：
    - 模型难以收敛（数据分布差异过大）
- 解决方法
  - 提出离线数据异质性度量与在线概率采样优化相结合
    - 为客户端选择提供一个强大且隐私保护的“语义感知”衡量标准
    - 将离线的语义价值评估，转化为实际训练中高效的、动态的客户端选择策略，以实现系统整体效用的最大化



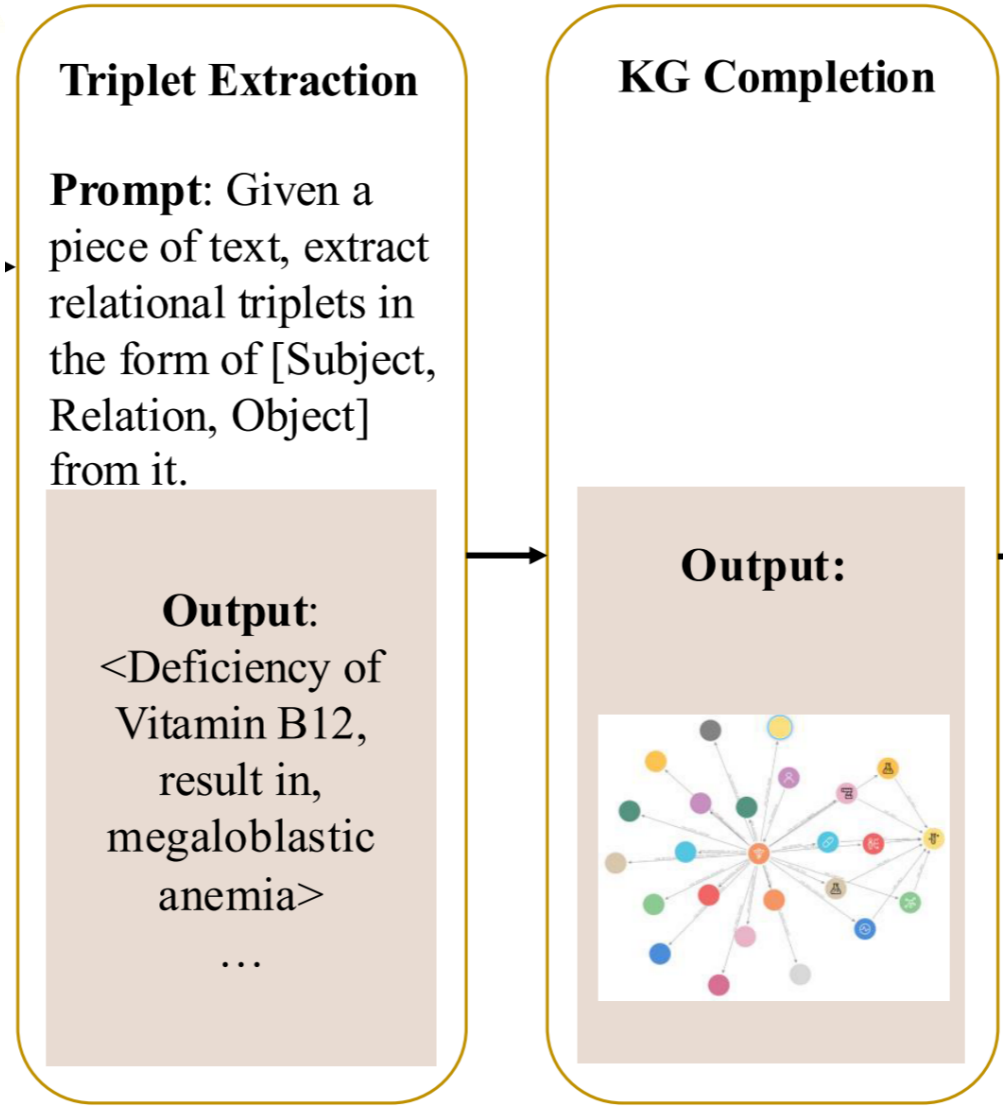
(a) Statistical Heterogeneity



(b) System and Statistical Heterogeneity

DDCS KG辅助客户端选择算法 离线阶段

- 构造对齐后的客户端知识图谱
  - 客户端文本分块：短文本/问答按条、长报告按段落等
  - 关系三元组抽取：用EDC对每块文本提取三元组，得到原始KG
  - KG补全：用GreenKGC（低维的KGE）对原始KG做补全以恢复可能缺失的边/关系，从而提高KG的完整性与可比性（减小噪声）
  - KG对齐：不同客户端构造的KG可能命名不一致，使用AutoAlign做实体与谓词对齐。完成对齐后，KG可跨客户端比较



DDCS KG辅助客户端选择算法 离线阶段

- 散度定义（divergence）
  - 对S个被选客户端集合S_t的散度定义

$$div(\mathcal{S}_t) = \sum_{\forall i,j \in \mathcal{S}_t}^{i \neq j} 1 - \frac{|g(\mathcal{D}_i) \cap g(\mathcal{D}_j)|}{|g(\mathcal{D}_i) \cup g(\mathcal{D}_j)|}$$

客户端i对齐后知识图谱的三元组集合

  - 基于Jaccard类似度，将对齐后三元组集合的重叠作为相似度，直接反映语义/事实层面的重合度
  - 该指标越大，代表双方数据语义差异越高；越小则表示知识重叠多，联合训练收益大，把 $\mathrm{div}(S_t)$ 用作约束以保证被选客户端的数据不要过于异质，从而加速收敛
- 客户端被选概率
  - 在联邦大模型微调中，服务器在每一轮通信中从N个客户端中独立采样S个客户端参与训练（有放回）

$$p_i \in [0,1], \quad \sum_{i=1}^{N} p_i = 1.$$

DDCS KG辅助客户端选择算法 离线阶段

- 客户端概率选择优化
  - 通过优化每个客户端被选中的概率p_i，在保证数据分布多样性（散度可控）的前提下，最小化总训练的wall-clock时间

最小化wall-clock时间

$$\min_{\mathbf{p}} \sum_{t=1}^{T} \max_{i \in \mathcal{S}^t} t_i$$

客户端 i 被选中的概率

$$s.t. \sum_{i=1}^{N} p_i = 1$$

第 t 轮被选中的客户端集合

$$\forall \mathcal{S}^t \in \mathcal{N}, |\mathcal{S}^t| = S$$

$$\forall \mathcal{S}^t \in \mathcal{N}, \mathbb{E}(div(\mathcal{S}^t)) \leq \epsilon$$

每轮期望最大时延：

$$\sum_{i=1}^{N} \left[\left(\sum_{j=1}^{i} p_j \right)^S - \left(\sum_{j=1}^{i-1} p_j \right)^S \right] t_i T$$

第 i 个客户端成为“最慢节点”的概率

总体目标：

$$\min_{p} \quad T \sum_{i=1}^{N} \left[\left(\sum_{j=1}^{i} p_j \right)^S - \left(\sum_{j=1}^{i-1} p_j \right)^S \right] t_i.$$

被选客户端的数据分布差异（由基于散度的KG度量）满足，使得系统异质性（延迟）与采样概率p_i的关系显式可计算

DDCS KG辅助客户端选择算法 在线阶段

- 在线抽样与无偏聚合
  - 服务器根据提前求得的概率分布p^*按有放回方式抽样得到$multiset\ S_t$，并把当前全局模型w_t下发给客户端
  - 每个客户端用 LoRA（更新 A, B）执行E个本地 epoch（或者E次局部更新），产生本地 LoRA 模块$A_{i,E}$，$B_{i,E}$
  - 服务器用调整后的加权聚合把这些 LoRA 模块聚合回全局，并对被抽中多次的客户端按抽中次数修正权重

$$w_{t+1} = w_t + \sum_{i \in S_t} \frac{d_i}{S\, p_i} \left(w_{t+1}^i - w_t \right).$$

其中$|S_t| = S$，p_i为该客户端在 multiset 中的抽样概率，t为轮数

DDCS 无偏聚合验证

- 全客户端参与聚合

$$\bar{\mathbf{w}}^{t+1} := \sum_{i=1}^{N} d_i \mathbf{w}_i^{t+1}.$$

- 其中$d_i = \frac{|D_i|}{|D|}$
- 论文的聚合方式

$$w_{t+1} = w_t + \sum_{i \in S_t} \frac{d_i}{S p_i} \left(w_{t+1}^i - w_t\right).$$

Algorithm 1: Federated LoRA with Probabilistic Client Selection

Input: Sampling probabilities $\mathbf{p} = \{p_i, \forall i \in \mathcal{N}\}$, initial model $\mathbf{w}^0$
Output: Final model parameter $\mathbf{w}^t$

for $t = 1$ *to* T **do**
  Server randomly samples a subset of clients $\mathcal{S}_t$ according to $\mathbf{p}$, and distribute global model $\mathbf{w}^t$ to the selected clients $\mathcal{S}_t$; **for** $i \in \mathcal{S}_t$ **do**
    LocalLoRAUpdate($\mathbf{w}^t$)
  $\mathbf{w}^{t+1} \leftarrow \mathbf{w}^t + \sum_{i \in \mathcal{S}_t} \frac{d_i}{Sp_i} \mathbf{B}_{i,E}^t \mathbf{A}_{i,E}^{t\,T}$
LocalLoRAUpdate($\mathbf{w}^t$): **for** *j=0,1,…,E-1* **do**
  $\mathbf{A}_{i,j+1}^t \leftarrow \mathbf{A}_{i,j}^t - \eta \Delta(\mathbf{A}_{i,j}^t)$;
  $\mathbf{B}_{i,j+1}^t \leftarrow \mathbf{B}_{i,j}^t - \eta \Delta(\mathbf{B}_{i,j}^t)$;

- 基于随机抽样S_t取期望

$$\mathbb{E}_{S_t}[w_{t+1}] = w_t + \sum_{i=1}^{N} \frac{d_i}{Sp_i} (Sp_i) \left(w_{t+1}^i - w_t\right) = w_t + \sum_{i=1}^{N} d_i \left(w_{t+1}^i - w_t\right). \longrightarrow \mathbb{E}_{\mathbb{S}^t}(\mathbf{w}^{t+1}) = \sum_{i=1}^{N} \frac{d_i}{Sp_i} Sp_i \mathbf{w}_i^{t+1} = \bar{\mathbf{w}}^{t+1}$$

实验设计　数据资源

- 数据集
  - 医学多项选择题回答(MedMCQA)
    - MedQA
    - MedMCQA
    - MMLU-Medical
  - 医疗报告摘要(MRS)
    - MIMIC-III
- 评价指标
  - MedMCQA指标：使用测试准确度（正确答案与问题总数的比率）
  - MRS指标：使用BLEU和Rouge-L作为句法度量生成的文本在语法上是否与自然文本相似

- BLEU：衡量机器生成文本的“精确度”。它重点关注“生成的词或短语是否出现在参考答案中”
- Rouge-L：衡量机器生成文本的“召回率”。它重点关注“参考答案中重要的词或短语是否被生成文本覆盖了”

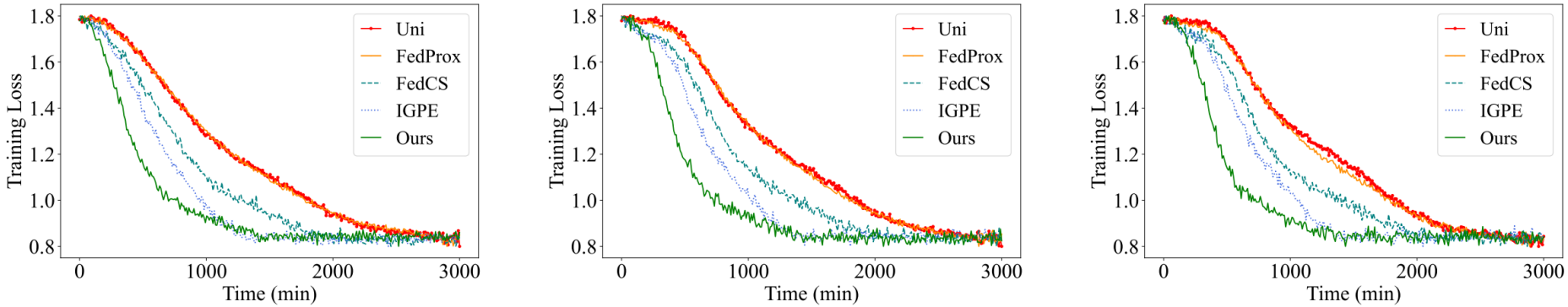
实验设计　数据资源

- 对比方法
  - Uniform Sampling (UNI)、FedProx、FedCS、IGPE
- 预训练模型选择
  - LLama2-7b、Mistral-7b
- 系统异构与统计异构的设置
  - IID：所有客户端数据集都是IID
  - Class：客户只拥有从所有主题中随机选择的数据。在没有进一步说明的情况下，将μ设置为50
  - Dirichlet：通过Dirichlet过程划分训练数据集。并且每个客户端保存来自不同主题的不同数量的数据，这些主题不均匀地分布

实验设计　对比实验

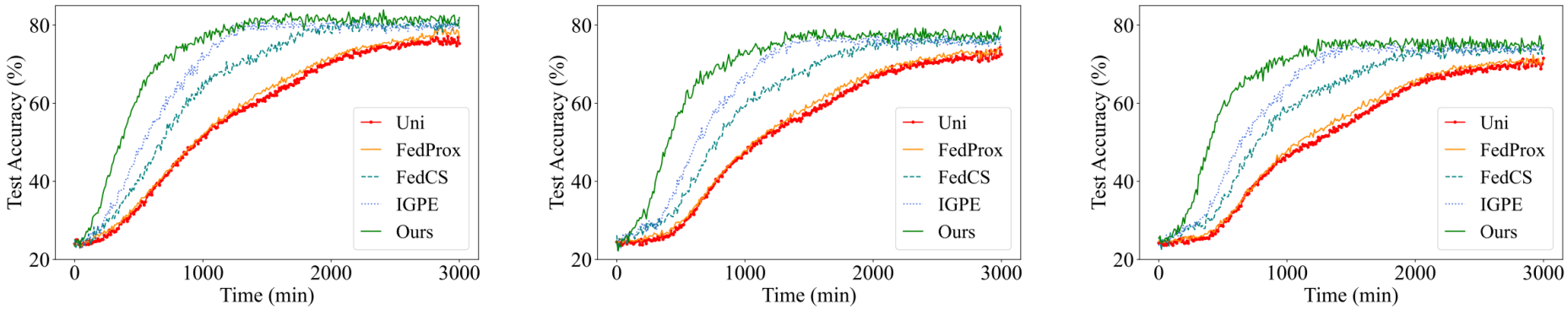
- 对于MedMCQA任务评估
  - 方案在微调的Llama2-7b模型中收敛速度和准确率都优于对比方法，方案能够很好地处理联合LLM微调中统计异构性和系统异构性共同存在的问题

收敛速度



(a) IID　(b) Class　(c) Dirichlet

准确率

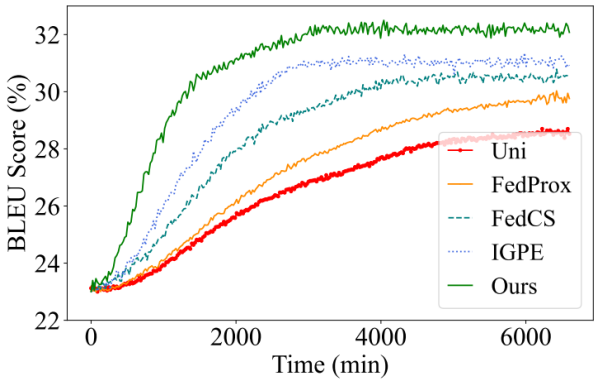


(a) IID　(b) Class　(c) Dirichlet

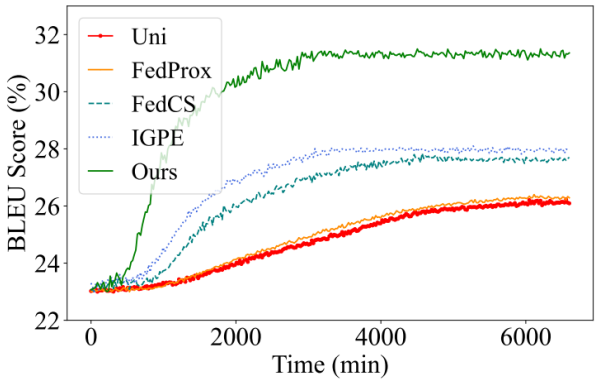
实验设计　对比实验

- 对于MRS任务以BLEU得分和Rouge-L得分评估
  - 方案在微调的Lllama2-7b模型生成的摘要文本的句法质量比其他基线收敛得快得多，当数据设置不是IID时，由方案微调的模型的效用仅受轻微影响

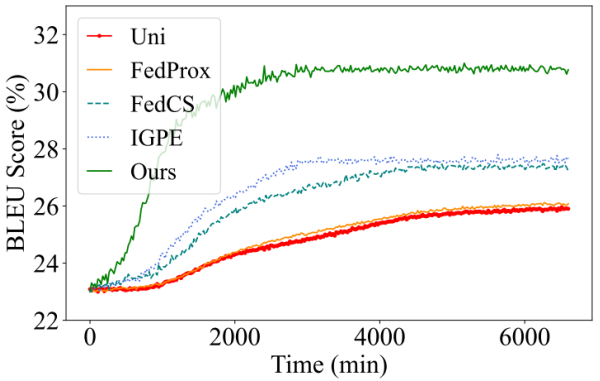
BLEU



(a) IID

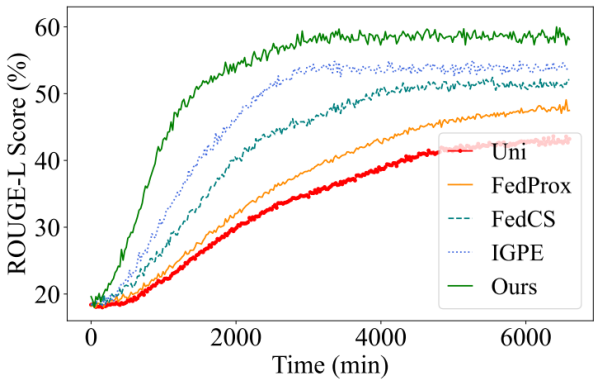


(b) Class

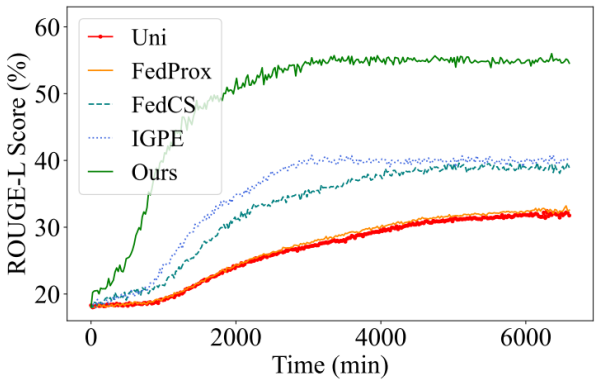


(c) Dirichlet

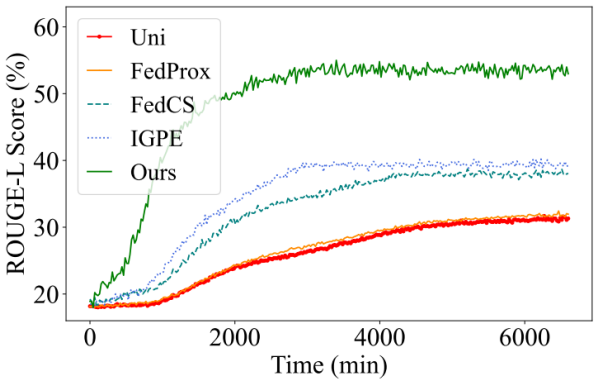
ROUGE-L



(a) IID



(b) Class



(c) Dirichlet

实验设计　对比实验

- **Mistral-7b模型的微调性能**
  - 对于MedMCQA应用，准确率提高了5.56%，收敛速度比基线快2.09倍
  - 对于MRS应用，显著加快了收敛速度，提高了任务性能

Data Setting	Baselines	Metrics	
		Accuracy	Latency
IID	*Uni*	75.84	2876
	FedProx	77.05	2786
	FedCS	79.58	2213
	IGPE	80.06	1457
	Ours	**81.28**	**1376**
Class	*Uni*	71.85	2863
	FedProx	73.15	2798
	FedCS	75.18	2267
	IGPE	76.06	1536
	Ours	**77.28**	**1423**
Dirichlet	*Uni*	69.72	2900
	FedProx	70.79	2826
	FedCS	73.07	2243
	IGPE	74.03	1512
	Ours	**75.28**	**1417**

MedMCQA

Data Setting	Baselines	Metrics			
		BLEU	ROUGE-L	BERTScore	Latency
IID	*Uni*	30.12	59.17	88.75	6327
	FedProx	31.55	60.08	89.48	6129
	FedCS	33.09	61.99	90.96	4860
	IGPE	33.76	62.23	91.64	3210
	Ours	**35.17**	**64.26**	**94.89**	**3017**
Class	*Uni*	26.52	52.95	79.37	6321
	FedProx	27.39	54.49	80.38	6172
	FedCS	28.16	55.85	81.75	4798
	IGPE	28.55	56.12	82.67	3256
	Ours	**32.77**	**59.26**	**87.89**	**3066**
Dirichlet	*Uni*	26.03	52.01	78.02	6410
	FedProx	26.11	52.11	79.25	6139
	FedCS	27.73	53.89	81.26	4892
	IGPE	27.82	54.04	81.51	3257
	Ours	**32.12**	**59.22**	**87.07**	**3089**

MRS

实验设计　消融与参数实验

- 评估KG completion
  - Variation-A：从方案中排除了KG补全，并且每个客户持有具有潜在缺失链接的原始知识图
- 评估KG alignment
  - Variation-B：从方案中排除了KG对齐，直接使用每对KG来估计客户数据的散度，而不对齐它们
- Variation-A的KG中缺少的链接和Variation-B的KG中语义相同但未对齐的节点

Schemes	Metrics	
	Accuracy	Latency
Variation-A	73.02	1922
Variation-B	72.45	2055
Ours	76.16	1416

算法总结 DDCS

- 算法贡献
  - 首次将知识图谱用于联邦LLM微调客户端选择，通过离线度量数据异构并联合优化系统延迟，实现更快收敛且提升模型精度
  - 通过修正权重聚合与带放回抽样策略，使全局更新在期望意义下无偏，保证理论收敛
  - 同时考虑了谁更慢与被选到的客户端数据是否均匀

- 算法不足
  - 知识图谱提取受文本噪声、命名歧义影响，可能导致关系错误或语义丢失
  - 客户端的语料或系统状态可能随时间变化，但知识图谱在离线阶段固定生成

LLM-driven Medical Report Generation via Communication-efficient Heterogeneous Federated Learning

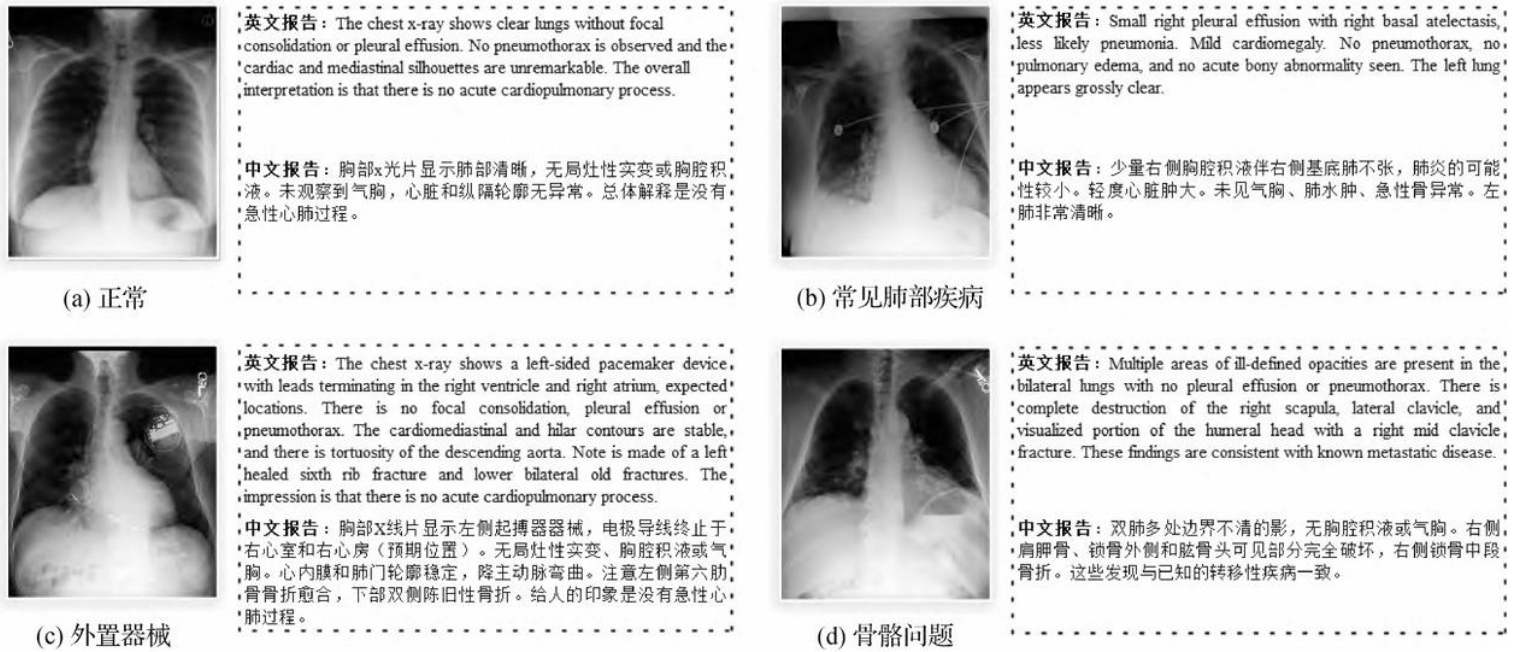
FedMRG TIPO

T	目标	在保持隐私的前提下利用多中心影像-报告对联合训练LLM
I	输入	LLaMA-2-7B-Chat作为基础大语言模型；MIMIC-CXR：276778张报告-图像、IU X-Ray：4168张图像、CheXpert+：65,240名患者的224,316张胸片
P	处理	1. 通信高效化（LoRA） 2. 图像端特征建模（HCP 模块） 3. 文本端语言建模（DMB 模块） 4. 联邦聚合与更新
O	输出	自动生成的医学报告；性能指标结果：在BLEU、ROUGE、CIDEr 等自然语言生成指标及临床一致性

P	问题	1. 联邦LLM调优中通信开销的基本挑战 2.FL情景下MRG的双重异质性：不同医疗中心的图像特征不同，以及不同的报告风格和术语偏好
C	条件	同步更新机制、数据分布异构（Non-IID）但任务一致
D	难点	1.如何降低联邦学习过程的通信开销 2.如何克服双重异质性挑战
L	水平	2025 SCI中科院1区Top

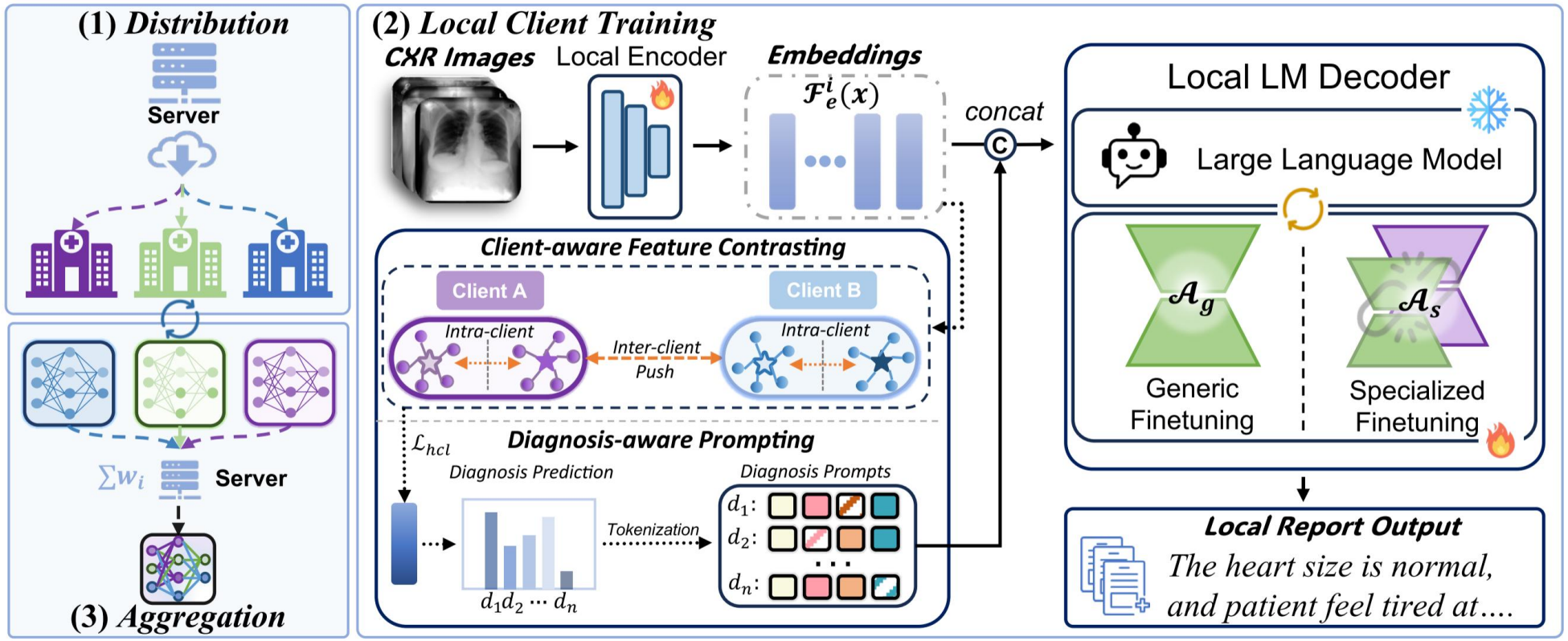
知识基础　医学报告生成（MRG）

- 医学报告生成（MRG）
  - 旨在从医学图像中自动创建叙事文本
  - 目标：识别临床异常和生成较长的报告
- 关键挑战
  - 医疗数据的有限可用性直接限制了生成报告的准确性，特别是对于数据饥渴的LLM驱动的MRG系统



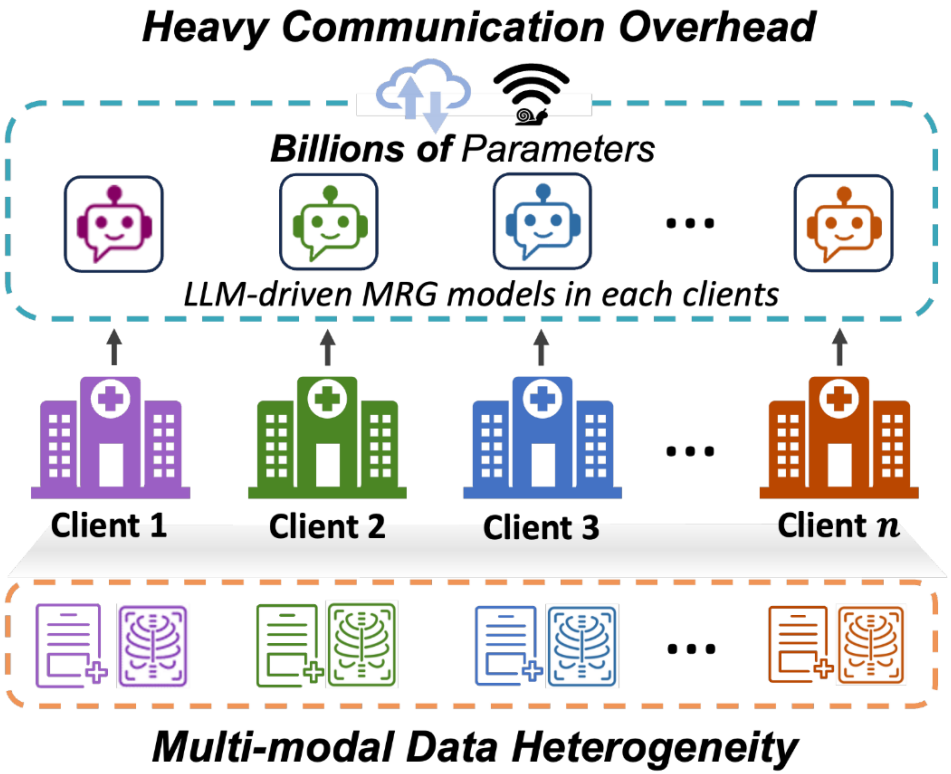
FedMRG 算法原理

- 算法原理图
  - 分发：服务器将初始化的模型分发给客户端
  - 本地训练：客户端使用层次对比与提示（HCP）与双适配器互增强（DMB）训练模型
  - 聚合：仅上传编码器参数和通用适配器，以减少通信开销，服务器将其聚合并重新分配



FedMRG 创新点分析

- 现有方法存在问题
  - LLM的大量参数使得使用传统的联邦算法进行训练在通信开销方面代价高昂
  - 在学习全局通用特征的同时保留客户端特定特征，特别是在处理跨中心的异构医学图像数据
- 解决方法
  - LoRA（低秩适配）的参数化：削减可训练参数量，降低通信开销，为LLM联邦训练奠定基础
  - 分层对比与诊断提示：缓解影像异构，注入临床先验，提升报告准确性
  - 双适配器互增强：捕获全局与本地文本模式，通过知识蒸馏双向增强，兼顾标准与个性化



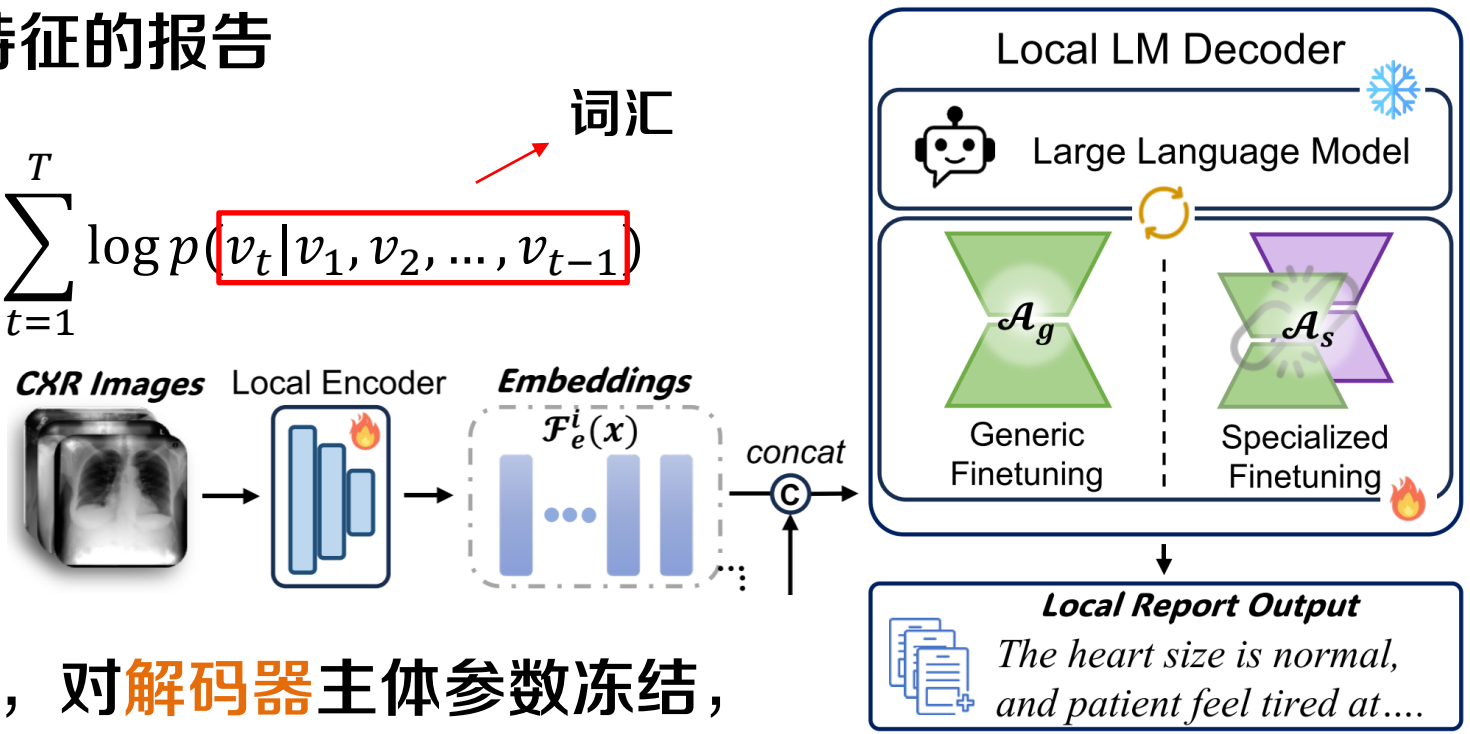
FedMRG LoRA（低秩适配）的参数化

- FedMRG从图像中提取视觉特征的视觉编码器和具有可训练组件的冻结LLM解码器，该组件生成基于视觉特征的报告

$$\mathcal{L}_{LM}(r', r) = -\sum_{t=1}^{T} \log p(v_t | v_1, v_2, \ldots, v_{t-1})$$

词汇

生成的文本与参考真实文本



- 在 LLM 解码器层使用 LoRA，对解码器主体参数冻结，仅传输编码器与通用适配器参数，采用 LoRA 使得解码器微调通信成本大幅下降

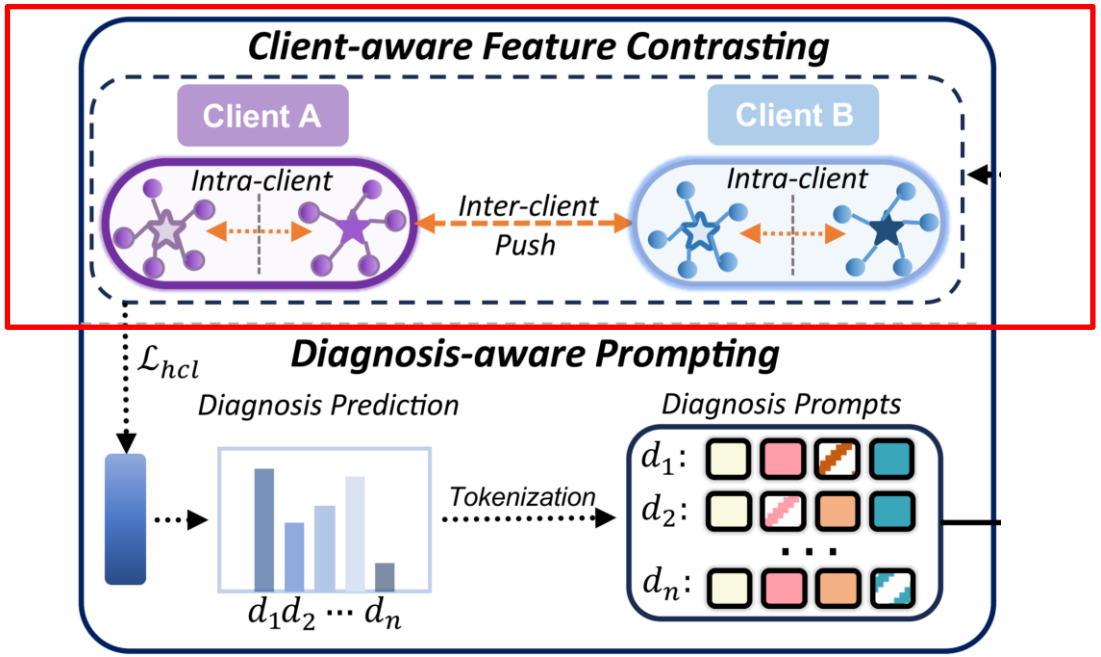
$$W'_m = W_m + \Delta\theta^b_m \Delta\theta^a_m$$

FedMRG 层次对比与提示（HCP）图像端

- 客户端感知特征对比
  - 在单个客户端上对局部正样本做对比，同时利用全局记忆库提供跨客户端负样本

$$\mathcal{L}_{hcl}^{i} = -\log \frac{\exp(f_{avg}^{i} \cdot f_{avg}^{j(i)} / \tau)}{\sum_{a \in A(i)} \exp(f_{avg}^{i} \cdot f_{avg}^{a} / \tau)},$$

  - 其中·符号为内积，τ为温度参数，$A(i) \equiv (J \cup I \cup M)/\{i, i'\}$，在本地层次与跨客户端层次都采集正/负样本以增强鲁棒性

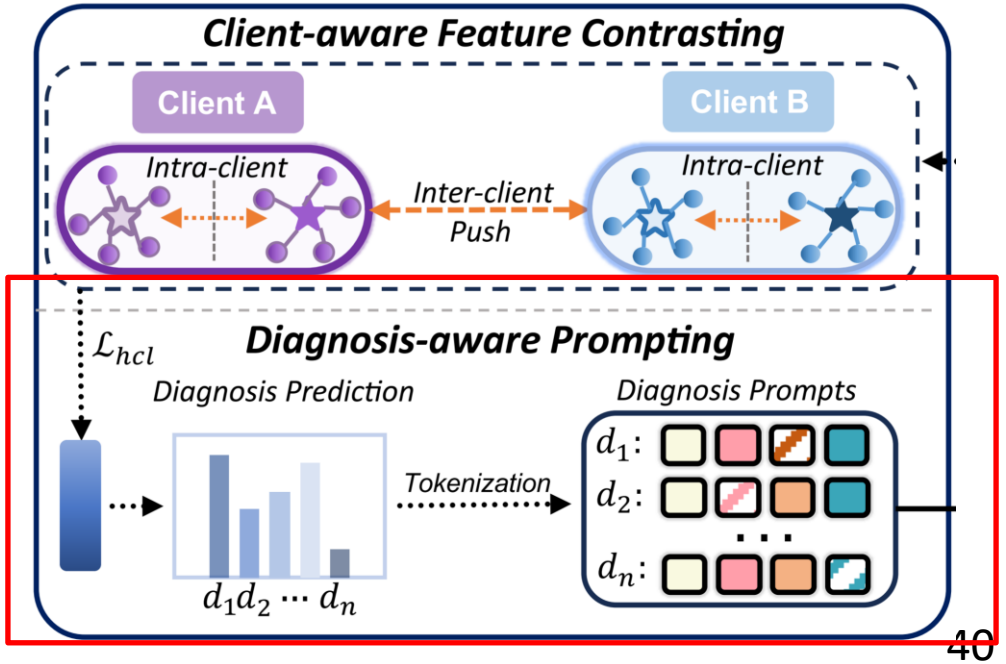


同时学习局部变体与全局可迁移特征

FedMRG 层次对比与提示（HCP）图像端

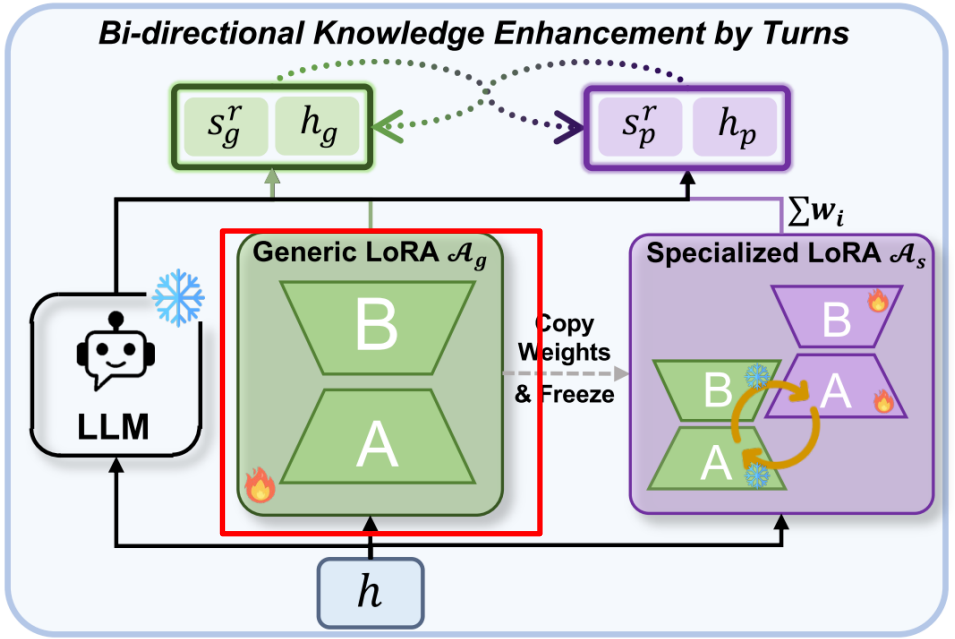
- 诊断感知提示
  - 为了捕获相同诊断下报告之间的相似性，将诊断预测转换为文本解码器的输入提示，为文本生成提供强大的临床指导
  - 疾病分类分支，诊断分支将14种胸部疾病预测转化为（空白、阳性、阴性、不确定）四类提示词，输入LLM指导生成，确保报告与影像诊断保持一致

HCP把视觉表征和诊断性提示结合，既保证诊断敏感信息被编码到视觉特征，又用显式提示引导 LLM 生成更临床相关的文本



FedMRG 双适配器相互增强（DMB）文本端

- 双适配器相互增强（DMB）
  - 通用适配器A_g：用于编码全局报告模式，同时最大限度地减少通信开销
  - 专用适配器A_s：用于保留特定于客户端的报告风格，而无需参数共享
- 通用适配器作为标准的LoRA模块实现，并通过通用调优进行优化，通过传统的联邦平均算法聚合全局知识

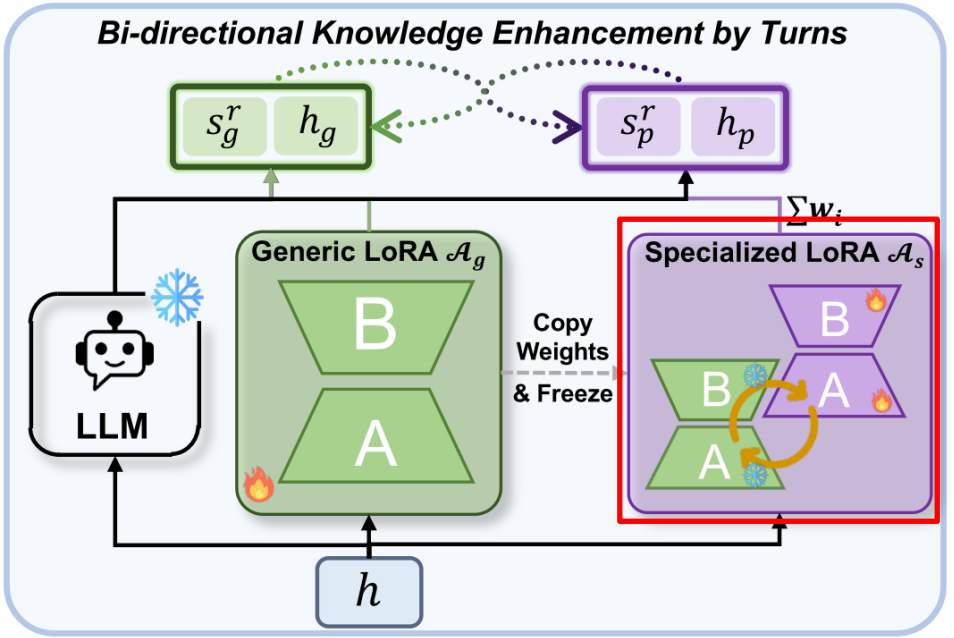


FedMRG 双适配器相互增强（DMB）文本端

- 专用适配器包括两个互补的LoRA模块，经过专门的微调训练
  - 第一个模块作为全局知识的储存库，在每个局部训练回合开始时继承通用适配器的权重，然后冻结
  - 第二个模块扩展了全局基础，以捕获特定于本地客户的细微差别，从而促进专业知识的获取

$$\mathcal{L}_{l2g} = \mathcal{L}_{cos}(h_s, h_g) + \mathcal{L}_{KL}(s_g^r || s_s^r),$$
$$\mathcal{L}_{g2l} = \mathcal{L}_{cos}(h_g, h_s) + \mathcal{L}_{KL}(s_s^r || s_g^r),$$

DMB 兼顾通信成本（只上传 A_g ）与本地个性化（A_s），双向蒸馏确保本地与全局知识互补，从而在保留个性化的同时提升全局模型质量



FedMRG 总体框架

- 初始化阶段
  - 服务器创建n个本地MRG模型并将其分发给参与的客户端，每个模型的解码器都配备了通用和专用适配器
- 客户端训练阶段
  - 优化模型：
    - 适配器优化，使用LLM进行基本报告生成功能，$\mathcal{L}_{l2g}$，$\mathcal{L}_{g2l}$用于适配器之间的双向知识转移
    - 编码器训练，利用分层对比和诊断感知提示
  - 完成本地训练后
    - 客户端只上传他们的编码器参数和通用适配器参数，以尽量减少通信开销

Algorithm 1 Pseudocode of FedMRG

Server Initialization:
for each client $i = 1, 2, \ldots, n$ **do**
  Initialize a MRG model with $\mathcal{F}_e$ and $\mathcal{F}_d$
  Freeze parameters of $\mathcal{F}_d$
  Insert $\mathcal{A}_s$ and $\mathcal{A}_g$ into $\mathcal{F}_d$
  Distribute initialized model to client i
end for
repeat
  Local Client Training:
  for each client $i = 1, 2, \ldots, n$ **in parallel do**
    for each mini-batch $j = 1, 2, \ldots, N^i$ **do**
      Compute $\mathcal{L}_1 := \alpha\mathcal{L}_{HCL} + \mathcal{L}_{CE} + \beta\mathcal{L}_{l2g} + \mathcal{L}_{LM}$
      Optimize $\mathcal{F}_e$ and each $\mathcal{A}_g$ via $\mathcal{L}_1$
      Compute $\mathcal{L}_2 := \alpha\mathcal{L}_{HCL} + \mathcal{L}_{CE} + \beta\mathcal{L}_{g2l} + \mathcal{L}_{LM}$
      Optimize $\mathcal{F}_e$ and each $\mathcal{A}_s$ via $\mathcal{L}_2$
      Upload $\mathcal{F}_e$ and each $\mathcal{A}_g$ and $\mathcal{A}_s$
    end for
  end for
  Server Aggregation and Distribution:
  Aggregate clients updated $\mathcal{F}_e$ and $\mathcal{A}_g$
  Distribute the aggregated components
until maximum communication rounds reached

FedMRG 总体框架

- 聚合阶段
  - 服务器对接收到的模型组件进行参数平均
- 分发阶段
  - 聚合的参数被传输回来以重新初始化客户端模型

Algorithm 1 Pseudocode of FedMRG

Server Initialization:
for each client $i = 1, 2, \ldots, n$ **do**
    Initialize a MRG model with $\mathcal{F}_e$ and $\mathcal{F}_d$
    Freeze parameters of $\mathcal{F}_d$
    Insert $\mathcal{A}_s$ and $\mathcal{A}_g$ into $\mathcal{F}_d$
    Distribute initialized model to client i
end for
repeat
    Local Client Training:
    for each client $i = 1, 2, \ldots, n$ **in parallel do**
        for each mini-batch $j = 1, 2, \ldots, N^i$ **do**
            Compute $\mathcal{L}_1 := \alpha\mathcal{L}_{HCL} + \mathcal{L}_{CE} + \beta\mathcal{L}_{l2g} + \mathcal{L}_{LM}$
            Optimize $\mathcal{F}_e$ and each $\mathcal{A}_g$ via $\mathcal{L}_1$
            Compute $\mathcal{L}_2 := \alpha\mathcal{L}_{HCL} + \mathcal{L}_{CE} + \beta\mathcal{L}_{g2l} + \mathcal{L}_{LM}$
            Optimize $\mathcal{F}_e$ and each $\mathcal{A}_s$ via $\mathcal{L}_2$
            Upload $\mathcal{F}_e$ and each $\mathcal{A}_g$ and $\mathcal{A}_s$
        end for
    end for
    Server Aggregation and Distribution:
    Aggregate clients updated $\mathcal{F}_e$ and $\mathcal{A}_g$
    Distribute the aggregated components
until maximum communication rounds reached

实验设计　数据资源

- 数据集
  - 胸部X射线数据集
    - MIMIC-CXR：经过标准化预处理的最大MRG数据集，包含276778张带有相应报告的图像
    - IU X-Ray：一个广泛采用的MRG评估数据集，包含4168张图像
    - CheXpert+：来自65,240名患者的224,316张胸片（包括正位和侧位）
- 实验设置
  - 客户端异构设计
    - FL-MRG (Random)：病人级别随机分配
    - FL-MRG (Clustering)：基于报告聚类的分配（强异质性）
  - 现实世界的联邦设置
    - 用 CheXpert+ 将真实机构数据划分以构建多源真实设置
- 评价指标
  - 自然语言生成指标（NLG）
    - BLEU（BL1-BL4）、CIDEr（CID）、ROUGEL（ROU）
  - 临床疗效指标（CE）
    - 基于示例的精度评分（PRE）、基于示例的F1评分（F1）

实验设计　对比实验

- 评估FedMRG在不同数据集上的表现（ MIMIC-CXR ）
  - LLM驱动的MRG模型始终优于传统的MRG方法
  - FedMRG与只做 LoRA / 只做 prompt 的基线比较，FedMRG的综合指标更好

Model	Comm.	FL-MRG (Cluster)								FL-MRG (Random)							
		BL1	BL2	BL3	BL4	ROU	F1	REC	PRE	BL1	BL2	BL3	BL4	ROU	F1	REC	PRE
Federated & MRG Baseline																	
Centralized	-	32.70	20.40	13.90	13.43	27.71	29.9	27.8	37.8	32.70	20.40	13.90	13.43	27.71	29.9	27.8	37.8
Transformer [28]	63M	27.05	17.00	11.43	8.16	26.54	16.1	13.8	23.3	27.68	17.48	11.86	8.54	26.82	19.0	16.0	28.3
R2Gen [28]	82M	26.93	16.80	11.33	8.15	25.94	17.3	14.7	25.4	27.95	17.62	11.90	8.54	26.73	20.8	17.2	32.1
WCL [37]	203M	37.28	23.02	15.31	10.87	26.60	32.0	29.6	41.0	32.54	18.63	12.24	8.99	26.13	27.8	31.3	37.5
DCL [33]	203M	36.96	22.59	15.10	10.86	26.75	28.4	26.2	36.3	37.11	23.46	15.42	11.22	27.08	29.6	27.1	38.7
LLM-Driven MRG																	
R2GenGPT [1]	51M	39.40	24.94	17.01	12.27	27.22	26.6	23.8	35.7	38.78	24.62	16.87	12.19	27.04	29.1	27.3	36.9
PromptMRG [7]	59M	38.23	24.05	16.43	11.79	26.70	30.7	26.4	43.2	39.20	24.87	17.04	12.26	27.14	32.2	28.5	43.3
Fed-AdaLoRA [55]	57M	39.75	24.90	16.85	12.03	27.10	29.9	27.2	38.7	40.31	25.42	17.30	12.40	27.30	30.8	28.4	39.0
FedPara [57]	76M	38.48	24.16	16.38	11.74	26.13	26.4	23.8	35.0	39.78	25.03	16.99	12.19	27.11	27.3	23.9	37.5
Fed-LoRA [52]	59M	39.30	24.68	16.75	12.01	26.82	25.1	21.8	35.2	40.13	25.50	17.45	**12.57**	**27.34**	30.0	28.0	38.0
Fed-Prefix [53]	103M	38.46	24.10	16.43	11.90	26.77	26.9	24.1	35.5	39.57	24.88	16.94	12.19	27.06	28.0	25.4	36.3
Fed-Prompt [54]	102M	38.27	23.93	16.21	11.65	26.54	28.4	25.7	37.1	39.11	24.73	16.92	12.21	27.11	29.4	26.3	38.7
Fed-Vera [56]	51M	38.03	23.89	16.22	11.64	26.58	25.4	23.1	33.8	39.10	24.63	16.76	12.12	27.03	27.0	24.0	36.9
FedDAT [24]	88M	35.29	21.95	14.92	10.69	25.44	17.7	15.1	25.9	38.69	24.41	16.56	11.84	26.63	29.1	26.1	38.8
FedMRG (Ours)	59M	**40.19**	**25.52**	**17.55**	**12.68**	**27.55**	**33.9**	**30.2**	**45.4**	**40.48**	**25.80**	**17.80**	12.36	27.33	**35.6**	**32.4**	**45.8**

实验设计　对比实验

- 使用完整的IU X-Ray数据集，未见领域泛化测试
  - FedMRG保持了其优越的性能，证明了其对MIMIC和IU X-Ray数据集之间的域转移的弹性

Model	Comm.	FL-MRG (Cluster)								FL-MRG (Random)							
		BL1	BL2	BL3	BL4	ROU	F1	REC	PRE	BL1	BL2	BL3	BL4	ROU	F1	REC	PRE
Federated & MRG Baseline																	
Centralized	-	39.56	22.39	13.42	8.55	28.85	15.2	15.2	15.8	39.56	22.39	13.42	8.55	28.85	15.2	15.2	15.8
Transformer [28]	63M	32.04	18.62	11.25	6.66	26.07	13.62	13.44	14.20	34.15	19.10	12.06	7.73	26.10	14.2	14.1	14.6
R2Gen [28]	82M	32.50	17.78	10.14	6.20	25.68	13.4	13.8	13.2	35.91	20.45	12.06	7.50	26.83	13.3	13.1	13.7
WCL [37]	203M	32.23	16.64	8.80	5.05	21.99	13.46	13.26	6.30	36.31	21.58	12.33	8.01	27.53	13.5	13.5	14.0
DCL [33]	203M	31.36	16.53	9.06	5.36	23.26	13.27	13.18	14.06	35.40	19.92	11.56	7.15	25.48	13.0	13.0	13.4
LLM-Driven MRG																	
R2GenGPT [1]	51M	38.68	22.22	13.52	8.73	26.43	13.19	12.80	14.34	39.71	23.20	14.11	9.06	26.56	13.0	12.7	14.0
PromptMRG [7]	59M	42.18	25.33	16.28	11.05	**30.61**	18.17	15.76	16.39	42.05	25.10	15.99	10.79	**29.99**	17.0	14.9	15.4
Fed-AdaLoRA [55]	57M	36.93	21.08	12.71	8.11	25.66	13.01	12.78	13.91	38.40	22.04	13.17	8.29	26.28	12.9	12.7	13.8
FedPara [57]	76M	37.40	21.17	12.45	7.76	25.65	11.71	11.49	12.36	38.10	21.82	13.06	8.26	26.27	12.4	12.1	13.3
Fed-LoRA [52]	59M	38.04	21.74	13.12	8.34	25.74	12.69	12.38	13.71	39.88	23.15	14.04	9.00	27.10	14.0	13.7	15.0
Fed-Prefix [53]	103M	37.68	21.63	13.19	8.53	25.82	13.07	12.76	14.08	37.85	21.62	12.98	8.24	26.17	12.5	12.3	13.4
Fed-Prompt [54]	102M	36.80	20.78	12.37	7.80	25.53	12.38	12.12	13.27	39.42	22.71	13.63	8.64	26.47	13.2	12.9	14.2
Fed-Vera [56]	51M	37.23	21.05	12.54	7.96	25.38	12.68	12.35	13.60	39.95	23.28	14.30	9.26	26.69	12.7	12.4	13.7
FedDAT [24]	88M	36.16	21.30	13.62	9.19	26.02	11.11	10.92	11.69	38.53	22.32	13.41	8.49	26.19	13.9	13.7	14.9
FedMRG(Ours)	59M	**43.55**	**26.38**	**17.30**	**11.95**	29.73	**19.69**	**17.22**	**17.86**	**42.99**	**25.70**	**16.45**	**11.15**	29.66	**18.1**	**16.3**	**16.6**

实验设计　对比实验

- 评估FedMRG在真实联邦条件下的性能
  - 跨多个客户集成了MIMIC-CXR和CheXpert+数据集，以模拟真正的跨机构异质性
  - 在内部验证(MIMIC-CXR和CheXpert+)和外部测试方面，FedMRG始终优于所有基准和最先进的方法

Model	Comm.	Internal Test (MIMIC-CXR & Chexpert+)								External Test (IU X-ray)							
		BL1	BL2	BL3	BL4	ROU	F1	REC	PRE	BL1	BL2	BL3	BL4	ROU	F1	REC	PRE
Federated & MRG Baseline																	
Centralized	-	21.39	13.45	9.20	6.49	**24.92**	21.76	18.72	27.46	33.93	19.71	12.41	8.11	28.46	13.23	13.09	13.58
Transformer [28]	63M	18.33	11.73	8.01	5.69	24.82	14.28	12.45	18.55	36.98	23.34	15.98	11.49	30.56	12.49	12.44	12.59
R2Gen [28]	82M	19.51	12.36	8.38	5.87	24.82	19.35	16.48	27.17	32.47	19.84	12.89	8.68	29.41	13.04	12.94	13.32
WCL [37]	203M	28.68	16.77	10.21	6.51	22.63	15.09	13.50	20.69	26.94	14.99	8.48	5.25	21.85	5.44	5.52	5.52
DCL [33]	203M	29.97	17.64	10.98	7.19	22.99	10.59	9.86	13.05	38.90	22.77	13.62	8.63	25.99	9.85	9.84	9.91
LLM-Driven MRG																	
R2GenGPT [1]	51M	30.16	17.46	11.16	7.48	23.18	13.85	12.14	18.31	37.69	24.51	17.56	13.07	33.18	9.96	9.93	10.05
PromptMRG [7]	59M	34.57	20.68	13.29	8.90	24.08	12.65	10.73	17.99	37.01	23.01	15.65	11.18	28.16	10.03	9.97	10.20
Fed-AdaLoRA [55]	57M	31.72	18.55	11.89	7.98	23.32	15.87	13.74	21.78	37.61	24.51	17.66	13.18	33.12	10.07	10.01	10.24
FedPara [57]	76M	31.10	18.37	11.85	8.02	23.66	16.32	14.01	22.82	37.54	24.16	17.16	12.60	32.56	10.04	9.97	10.23
Fed-LoRA [52]	59M	**35.29**	20.38	12.79	8.48	23.54	19.96	17.06	28.14	36.75	23.60	16.50	11.98	30.22	10.50	10.93	12.35
Fed-Prefix [53]	103M	32.92	19.55	12.52	8.45	23.72	19.08	16.39	26.43	39.61	25.64	18.06	13.23	32.16	10.42	10.28	10.77
Fed-Prompt [54]	102M	34.10	20.79	13.56	**9.30**	24.75	17.71	15.46	24.10	**43.80**	27.95	19.53	14.19	29.81	10.21	10.11	10.50
Fed-Vera [56]	51M	32.60	19.18	12.43	8.42	23.56	14.88	13.23	19.39	39.01	25.90	18.90	14.27	**33.73**	9.96	9.93	10.03
FedDAT [24]	88M	33.01	19.27	12.26	8.09	23.58	21.17	19.01	28.18	39.11	24.86	17.54	12.88	32.43	10.58	10.36	11.18
FedMRG (Ours)	59M	**35.29**	**21.14**	**13.98**	8.90	24.34	**22.61**	**20.05**	**29.76**	42.49	**28.62**	**20.55**	**14.92**	33.57	**15.37**	**15.42**	**16.18**

实验设计　消融与参数实验

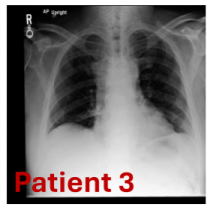
- 评估FedMRG的不同模块在不同数据集上的表现
  - 移除专用适配器、诊断提示与蒸馏损失等性能都有所下降，证实各模块协同贡献
  - 各模块都有可量化贡献

Model	BL1	BL2	BL3	BL4	ROU	F1	REC	PRE
FedMRG	**40.2**	**25.5**	**17.6**	**12.7**	**27.6**	**33.9**	**30.2**	**45.4**
· **w/o** $\mathcal{A}_s$	38.7	24.2	16.4	11.6	26.7	32.0	29.1	41.8
· **w/o** $\mathcal{L}_{hcl}$	37.9	23.7	16.1	11.5	26.7	30.8	27.5	41.3
· **w/o** p	38.2	24.1	16.4	11.8	26.8	29.3	26.3	38.5
· **w/o** $\mathcal{L}_{g2l}$	39.6	24.9	17.0	12.2	27.2	31.5	27.2	44.2
· **w/o** $\mathcal{L}_{l2g}$	39.6	24.9	16.9	12.2	27.1	30.8	26.3	43.9

实验设计　对比实验

- 评估基线和FedMRG的定性实例
  - 在不同案例报告的比较中，基线模型往往产生更笼统的预测，缺乏具体的描述性要素
  - FedMRG在多次扫描中显示出更强的连续性，注意到患者特有的发现

蓝色字体：模型正确再现真实报告
红色字体：模型能生成针对特定图像差异的个性化内容

	Ground-Truth	Baseline	FedMRG
 Patient 1  Patient 2	**GT1:** Frontal and lateral views of the chest were obtained. Rounded calcified nodule in the region of the posterior right lung base is seen and represents calcified granuloma on cts dating back to likely secondary to prior granulomatous disease. Previously seen lymph node and right hilar lymph nodes are better on ct. No focal consolidation is seen. There is no pleural effusion or pneumothorax. Cardiac and mediastinal silhouettes are stable with possible slight decrease in right paratracheal prominence. **GT2:** In comparison with the study of there is no evidence of pneumothorax. Continued low lung volumes with substantial mass in the right paratracheal region.	**Pred1:** The lungs are clear without focal consolidation. No pleural effusion or pneumothorax is seen. The cardiac and mediastinal silhouettes are unremarkable. **Pred2:** As compared to the previous radiograph there is no relevant change. Borderline size of the cardiac silhouette without pulmonary edema. No pleural effusions. no pneumonia. No pneumothorax.	**Pred1:** AP upright and lateral views of the chest provided. Midline sternotomy wires and mediastinal clips are again noted. There is no focal consolidation effusion or pneumothorax. The cardio and mediastinal silhouette is normal. Imaged osseous structures are intact. No free air below the right hemidiaphragm is seen. **Pred2:** As compared to the previous radiograph, there is no relevant change. The cardiac mediastinal and hilar contours appear stable. There is no pleural effusion or pneumothorax. The lungs appear clear.
 Patient 3  Patient 3	**GT3:** The lungs are low in volume. There is no focal airspace consolidation to suggest pneumonia. No pleural effusions, pulmonary edema, or pneumothorax. A calcified granuloma is present below the medial aspect of the right hemidiaphragm. The inferior sternotomy wire is fractured but unchanged. Post-CABG changes noted. **GT4 (Follow-up report):** Calcified pulmonary nodule in posterior right lung base, unchanged from priors. Enlarged right hilar and mediastinal lymph nodes causing opacity at right hilum and fullness in right paratracheal region. No pleural effusion, pneumothorax or focal consolidation. Post sternotomy changes with intact CABG wires. Cardiac silhouette normal in size.	**Pred3:** Pa and lateral views of the chest provided. Midline sternotomy wires and mediastinal clips are again noted. There is no focal consolidation effusion or pneumothorax. Imaged osseous structures are intact. **Pred4:** As compared to the previous radiograph there is no relevant change. Normal lung volumes. Normal size of the cardiac silhouette. Normal hilar and mediastinal structures. no pneumonia no pulmonary edema.	**Pred3:** The pulmonary vasculature is normal. Apart from subsegmental atelectasis in the left lung base. The lungs are clear. No pleural effusion or pneumothorax is seen. There are no acute osseous abnormalities. **Pred4:** Pa and lateral chest views were obtained with patient in upright position. The same holds for the previously described left - sided internal jugular approach central venous line. No pneumothorax has developed. The pulmonary vasculature is not congested. No signs of acute or chronic parenchymal infiltrates are present and the lateral and posterior pleural sinuses are free. No pneumothorax in the apical area.

算法总结 FedMRG

- 算法贡献
  - FedMRG首次实现大模型联邦医学报告生成，兼顾隐私、效率与质量，实验表明框架在多中心数据下生成报告临床准确且风格一致
  - 通信高效低秩适配：在 LLM 中采用低秩适配（LoRA）技术，仅训练少量适配器参数，显著减少通信和显存开销
  - 分层对比提示：通过层次对比学习（HCL）提升视觉表示的鲁棒性
  - 双适配器互增强：在客户端个性化与服务器端全球化之间建立了一个正向的增强循环
- 算法不足
  - 当前异构模拟仍基于公开数据集，真实临床疾病谱、患者人口学差异更复杂
  - 框架假设客户端同步在线，实际医院实际不符

特点总结与未来展望

特点总结与工作展望

- 特点总结
  - DDCS
    - 利用 KG 表示文本数据的语义与关系层次，能跨领域衡量客户端间的语义相似度，比传统基于样本均值/方差或梯度差的非IID度量更鲁棒、更语义相关
    - 构建 + 对齐知识图需要额外计算资源。若客户端数量大或文本规模大，离线阶段耗时明显
  - FedMRG
    - 采用 LoRA 低秩适配，仅上传小规模参数，通信量显著降低；支持资源受限场景
    - 通过双适配器互蒸馏机制实现全局知识与本地语言风格的平衡
    - 默认所有客户端同步参与聚合，不适合真实异步或断网场景
- 工作展望
  - 扩展至图像-文本等多模态数据，利用视觉/文本联合嵌入构建跨模态 KG
  - 设计在时延波动下的鲁棒采样与异步聚合机制，使算法适应移动端与 IoT 场景

参考文献

[1] B. Zhang, D. Wang, Y. Zhu and Z. Han, "Data Divergence-aware Client Selection via Knowledge Graph for Federated LLM Fine-tuning," IEEE Transactions on Mobile Computing, 2025.

[2] Che H, Jin H, Gu Z, et al. Llm-driven medical report generation via communication-efficient heterogeneous federated learning[J]. IEEE Transactions on Medical Imaging, 2025.

道德经

知人者智，自知者明。胜人者有力，自胜者强。知足者富。强行者有志。不失其所者久。死而不亡者，寿。

谢谢！