arXiv:2602.09306v1 [cs.LG] 10 Feb 2026

利用大语言模型赋能对比联邦序列推荐

Thi Minh Chau Nguyen、Minh Hieu Nguyen、Duc Anh Nguyen、Xuan Huong Tran、Thanh Trung Huynh 和 QuocViet Hung Nguyen

未提供所属机构

摘要。联邦序列推荐（FedSeqRec）旨在保护用户数据去中心化存储的同时进行下一物品预测，但各设备上存储的交互日志往往零散、含噪且同质，因而模型质量常受限制。许多现有方法试图通过手动数据增强或增加服务器端约束来弥补这一问题，但这些策略要么只能带来有限的语义多样性，要么会增加系统开销。为应对这些挑战，我们提出**LUMOS**，一种参数隔离的 FedSeqRec 架构，将大语言模型（大语言模型）作为本地语义生成器加以集成。LUMOS 不共享梯度或辅助参数，而是通过设备端大语言模型，在保护隐私的前提下，从每位用户的历史记录中构建三种互补的序列变体：（i）面向未来的轨迹，用于推断合理的行为延续；（ii）语义等价的改写，在保留用户意图的同时丰富交互模式；（iii）偏好不一致的反事实样本，作为信息丰富的负样本。这些合成序列通过三视图对比优化方案，在联邦主干网络中进行联合编码，从而在不暴露敏感信息的情况下实现更丰富的表征学习。在三个公开基准上的实验结果表明，与有竞争力的集中式和联邦基线相比，LUMOS 在 HR@20 和 NDCG@20 上均取得稳定提升。此外，即使没有专门的服务器端保护模块，利用语义基础充分的正向信号和反事实负样本，也能提升模型在含噪和对抗环境下的鲁棒性。总而言之，本研究展现了由大语言模型驱动的语义生成作为推进隐私保护联邦推荐新范式的潜力。

关键词： 联邦学习 · 序列推荐 · 大语言模型 · 对比学习 · 行为数据增强

1 引言

电子商务、新闻推送和在线媒体平台等个性化服务越来越依赖能够理解用户兴趣如何随时间变化的模型 [29]。序列推荐系统（SeqRecs）正是为此而设计的：它们不再将交互视为无序历史，而是明确

对有序的用户–物品序列进行建模，以推断用户可能会与之交互的下一个物品 [6,17]。早期研究通过纳入简单的时间依赖关系，扩展了传统的协同过滤 [7,8,31]; 之后的方法则采用神经网络架构来捕捉复杂的序列模式。代表性示例包括基于马尔可夫链的模型 [54], 、GRU4Rec 等循环架构 [15], 、SASRec 等基于自注意力的方法 [17] 以及 Cloze 风格的序列模型 [59]。近来，无 ID 序列推荐系统利用 BERT 等大型预训练语言模型（PLM） [2] 和更大型的大语言模型（LLM） [26,88,91,44] ，仅根据文本信号对交互序列进行编码。

然而，大多数此类序列模型都假设所有用户交互日志均被集中收集和存储。联邦学习（FL）提供了一种自然的替代方案：在不暴露原始数据的情况下，跨分布式客户端协同训练模型 [77]。将 FL 与序列推荐相结合后，便得到联邦序列推荐（FedSeqRec）：每个客户端保留自己的交互历史，并通过贡献本地梯度或预测参与全局模型更新 [19,86]。我们将采用这一范式的具体系统称为联邦序列推荐系统（FedSRS）。

尽管取得了令人鼓舞的进展，当前 FedSRS 的设计在有效利用用户本地序列方面仍有局限。首先，每个客户端可用的数据量通常很小，且高度稀疏、噪声较多，冷启动用户或低频用户尤其如此。在这种情况下训练强大的序列编码器，可能导致表示不稳定、泛化能力较差。其次，尽管对比学习已被证明是集中式推荐中有效的自监督信号 [78,80,16], ，但要将这些技术直接移植到 FedSeqRec 并非易事。典型的基于数据的增强（如随机裁剪、打乱、物品掩码）或基于表示的噪声注入，都是在集中式语料库规模庞大的假设下设计的 [73,79,69,74,71,65,81]。将这些启发式方法应用于较短的客户端特定序列时，可能会破坏用户行为的语义结构，或产生较弱且几乎冗余的视图。第三，在联邦学习中维护强有力的隐私保证会限制对跨客户端信息的访问，因此难以构建多样化的对比样本对，同时避免泄露敏感内容或传输模型参数 [67,24,23]。

与此同时，大语言模型在交互以自然语言表示时，展现出理解并生成连贯行为轨迹的强大能力 [26,88,91]。作为一种行为先验，大语言模型可以推断合理的未来、改述等价的偏好模式，或根据部分序列提出逼真的反事实样本。这引出了一个核心问题：我们能否利用客户端侧大语言模型为 *FedSeqRec* 合成更丰富的训练信号，同时不改变联邦学习协议或削弱隐私保证？

本文通过提出 LUMOS 来回答这一问题。LUMOS 是一个将大语言模型驱动的行为合成与多视图对比学习相结合的联邦序列推荐框架。在 LUMOS 中，每个客户端都将大语言模型视为本地行为生成器。以用户观测到的交互历史为条件，大语言模型会生成辅助序列，这些序列（i）在保持用户

意图的同时改变表层模式，或（ii）构成与用户真实偏好相冲突、难以区分但仍符合现实的负样本。这些合成序列绝不会上传至服务器；相反，它们会作为额外视图，用于完全在设备端优化的对比目标。全局服务器仅按照标准联邦学习协议聚合模型更新。不同于以往探索抵御对抗攻击的 FedSRS 工作 [36,66,55,83],，我们关注的是在良性联邦环境中提升表示质量和数据效率，并将大语言模型用作语义增强器，而非攻击或防御机制。我们在三个真实世界基准测试上评估了 LUMOS。实验结果表明，与集中式和联邦基线相比，LUMOS 始终能提升 top-K 准确率，尤其是在客户端历史记录稀疏且较短的情况下。消融研究进一步验证了大语言模型生成的视图和多视图对比目标的作用。

总而言之，本研究的贡献如下：

– 我们提出了 LUMOS，这是一个联邦序列推荐框架，利用大语言模型作为设备端行为生成器，为 FedSeqRec 丰富自监督信号。–我们提出了一种保护隐私的多视图对比学习方案，在每个客户端联合利用真实序列和大语言模型合成序列，同时不传输模型参数、嵌入或合成数据。– 我们基于多种广泛使用的序列推荐骨干模型实现了 LUMOS，并在三个公开数据集上进行了大量实验，结果表明其性能始终优于最先进的联邦基线。

2 相关工作

序列推荐系统。 现代在线平台（例如购物、阅读和媒体消费）中的用户行为往往会形成时间序列，揭示不断变化的用户兴趣。早期的协同过滤推荐系统 [7,8,31] 将交互视为无序集合，因此难以建模时间动态。为解决这一问题，Rendle等人 [54]提出了首批基于马尔可夫链的序列模型之一。随着深度神经网络的兴起，序列推荐迅速采用循环架构 [6], ，例如 GRU4Rec [15], ，随后又采用 SASRec [17], 等基于注意力的结构，展现出捕捉长程依赖的强大能力。Sun等人 [59] 通过引入一种类似 Cloze 的目标函数拓展了这一研究方向，从而实现双向上下文建模。大语言模型（LLM）的出现为序列推荐开辟了新的方向。无ID方法不再仅依赖物品 ID，而是利用文本语义和自然语言建模来表示用户历史记录。后续由大语言模型驱动的模型 [91] 表明，包括 BERT [2] 及更强大的大语言模型变体在内的大型预训练语言模型，无需专门的 ID嵌入，也能有效编码偏好序列。本研究基于这些见解，将大语言模型视为辅助行为生成器，以增强联邦环境中的序列建模。

联邦推荐系统。 联邦推荐旨在协调个性化需求与数据保护要求 [20,77,22]。最早的框架，例如 Ammad等人 [1], 提出的联邦协同过滤模型，证明了在不共享原始交互数据的情况下进行联合推荐是可行的。后续研究探索了更优的优化策略 [51,84,72],、更强的隐私保护 [21,52,85,87,89], 、鲁棒性考量 [36,66], 以及数据管理 [47,42,34,33,62,30,10]。

用于推荐的大语言模型和联邦大语言模型。 近年来，大语言模型已通过多种方式融入推荐系统。Lyu 等人 [25] 提示大语言模型改写物品描述，从而提升集中式推荐的质量。Wu 等人 [70] 以及 Shehmir 和 Kashef [58] 综述了 LLM4Rec 领域，涵盖判别式与生成式用途、适配策略（微调与提示），以及可扩展性、延迟和隐私等挑战。Tang 等人 [60]进一步指出，单个大语言模型可以充当领域无关推荐器。与此同时，联邦大语言模型学习（FedLLM）已发展为一个独立领域：Yao 等人 [75]回顾了在异构约束下训练大语言模型的技术，Ye 等人 [76] 则推出 FedLLM-基准，用于在真实的用户划分数据集上进行基准测试。我们的工作与 LLM4Rec 和 FedLLM 均有关联，但大语言模型的使用方式不同。我们不对大语言模型进行联邦化 [75,76] ，也不将其仅用作文本编码器或集中式推荐器 [25,70,58,60], ，而是假设每个客户端本地都可使用一个预训练大语言模型，并将其视为语义行为生成器。给定用户的历史记录后，大语言模型会生成辅助且符合行为规律的交互序列，我们将其用作联邦序列推荐的额外对比视图。据我们所知，这是首项在 FedSeqRec 中利用大语言模型生成的交互序列作为设备端对比视图的工作 [90,11,3,41,39,64,12]。**联邦推荐中的对比学习**。联邦推荐系统中的对比学习仍未得到充分探索。FedCL [67] 通过将模型参数发送至服务器来引入对比目标，并依靠本地差分隐私 [27] 缓解信息泄露。后续方法 [24,23] 遵循类似原则，通过交换嵌入或统计量开展协作，这会削弱联邦训练的隐私保证。与此同时，UNION [82] 将对比思想用于检测非定向攻击，而非在良性设置下提升表示质量。我们的工作则通过严格遵守联邦学习约束与之不同。相反，客户端依靠大语言模型生成的行为视图，通过多视图对比学习丰富训练过程。这既保护了隐私，又显著提升了数据稀缺的联邦环境中的序列表示 [61,5,45,9,32,53,35,63]。

3 模型与问题形式化

问题设定。 我们研究联邦序列推荐，其中一组用户 $U = \{u_1, \ldots, u_N\}$ 与全局物品语料库 $V = \{v_1, \ldots, v_M\}$进行交互。每位用户 $u \in U$ 都持有一个私有的、按时间排序的交互序列

$$S_u = [v_{u,1}, v_{u,2}, \ldots, v_{u,T_u}],$$

该序列完全保留在客户端设备上。目标是学习一个序列推荐模型 $f_\theta(\cdot)$，用于预测下一物品偏好分布：

$$\hat{y}_u = f_\theta(S_u),$$

据此，候选项中的下一物品按排名$v_{u,\ T_u+1}$ 排列于 V中。

联邦优化协议。我们采用标准的同步联邦学习协议。在每一轮 t中，服务器持有全局模型 $\theta^{(t)}$，并将其广播给抽样选出的用户子集 $\mathcal{U}_t \subseteq U$。每个客户端 $u \in \mathcal{U}_t$ 使用 $\theta^{(t)}$ 初始化本地模型，并在其私有数据 S_u上运行 E 轮随机梯度下降，生成更新后的模型参数 $\theta_u^{(t)}$。随后，服务器使用 FedAvg 规则聚合这些更新：

$$\theta^{(t+1)} = \frac{1}{|\mathcal{U}_t|} \sum_{u \in \mathcal{U}_t} \theta_u^{(t)}.$$

在整个过程中，原始交互序列、LLM 提示词和大语言模型生成的文本均保留在设备端；仅交换模型参数。

序列推荐主干网络。遵循序列推荐领域的先前工作 [15,17], ，我们假设 f_θ（·）由 GRU4Rec [15] 或 SASRec [17]等神经序列编码器实例化。给定一个序列 S_u，编码器会生成用户表示 $h_u = f_\theta$（S_u），随后将其投影为物品得分向量。用户 u 的标准下一物品预测损失为：

$$\mathcal{L}_u^{\mathrm{Rec}} = -\log p_\theta\big(v_{u,T_u+1} \mid S_u\big),$$

其中， $p_\theta(\cdot)$ 表示对 V 的预测概率分布。

客户端数据稀缺。 在联邦设置中，每个客户端通常只能观察到短且稀疏的交互历史，尤其是冷启动用户或低频用户。这使得使用掩码、裁剪或打乱等传统增强方法构造有信息量的自监督信号变得困难，因为这些方法假设序列较长且行为上下文更丰富。因此，纯粹的本地训练可能导致用户嵌入不稳定。

LLM 生成的行为视图。 为在不共享数据的情况下丰富本地监督，每个客户端都可以将预训练大语言模型（LLM）作为黑盒生成器使用。客户端以原始序列 S_u 为条件，私下向 LLM 发起查询，以获取若干合成行为变体：

$$S_u^{\mathrm{F}}, \quad S_u^{\mathrm{P}}, \quad S_u^{\mathrm{N}},$$

从较高层面看，这些变体分别对应用户行为的预测未来、保留意图的改述和反事实负面视图。这些序列绝不会传输到服务器；它们仅用于在客户端构造更丰富的训练信号。同一骨干网络 $f_\theta(\cdot)$ 将这些序列编码为表示：

$$h_u^{\mathrm{F}} = f_\theta(S_u^{\mathrm{F}}), \quad h_u^{\mathrm{P}} = f_\theta(S_u^{\mathrm{P}}), \quad h_u^{\mathrm{N}} = f_\theta(S_u^{\mathrm{N}}).$$

客户端目标与全局目标。在本地训练期间，每个参与的客户端 u 都会优化一个复合目标，该目标结合了监督预测与三视图对比学习：

$$\mathcal{L}_u = \mathcal{L}_u^{\mathrm{Rec}} + \lambda_{\mathrm{CL}}\, \mathcal{L}_u^{\mathrm{CL}},$$

其中， $\mathcal{L}_u^{\mathrm{CL}}$是第 4 节中定义的三视图对比损失， λ_{CL} 用于控制其相对于监督项的权重。

联邦训练目标如下：

$$\min_{\theta} \sum_{u \in U} \mathbb{E}\big[\mathcal{L}_u(\theta)\big],$$

其中，期望是对随机客户端采样 $\mathcal{U}_t$ 以及大语言模型生成和优化过程中的随机性取值。在这一表述下，服务器仅通过聚合模型更新来优化 θ ，而所有原始和合成的行为数据都严格保留在本地。

4 方法

本节介绍 LUMOS，这是一种通过大语言模型生成的行为视图丰富本地训练的联邦序列推荐框架。LUMOS 的整体架构如图1所示。其核心思想是使用预训练大语言模型作为本地行为生成器，以弥补每个客户端严重的数据稀疏性，然后将这些合成序列整合到三视图对比学习目标中。联邦学习协议本身仍采用标准的 FedAvg；所有创新都在于每个客户端如何在本地优化过程中构建并利用额外的语义视图。

4.1 Fed 化联邦序列推荐设置

我们采用第 3 节所述的联邦设置。每个用户 u 拥有一个私有序列 $S_u = [v_{u,1}, \ldots, v_{u,T_u}]$，共享骨干网络 $f_\theta(\cdot)$（例如 GRU4Rec [15] 或 SASRec [17]）将 S_u 编码为表示 $h_u = f_\theta(S_u)$。下一物品预测损失为：

$$\mathcal{L}_u^{\mathrm{Rec}} = -\log p_\theta\big(v_{u,T_u+1} \mid S_u\big),$$

模型更新通过 FedAvg [77]在服务器上进行聚合。原始序列和合成序列绝不会离开设备。

4.2 大语言模型驱动的行为增强

在标准 FedSeqRec 中，每个客户端通常只有少量交互，这限制了掩码或随机重排等传统增强方法的有效性。为缓解这一问题，LUMOS 假设每个客户端都可以在本地查询预训练大语言模型（或通过隐私保护 API 查询），并利用它生成以 S_u为条件的额外行为视图。我们考虑以下三种视图。

预测性未来视图。 该视图会对用户最后一次记录的交互之后可能出现的未来行为进行推断。给定序列 S_u，客户端会向大语言模型发出请求，生成一段简短延续，以呈现同一用户可能接下来选择的物品，从而构建一段面向未来的扩展序列。

$$S_u^{\mathrm{F}} = [v'_{u,T_u+1}, v'_{u,T_u+2}, \ldots].$$

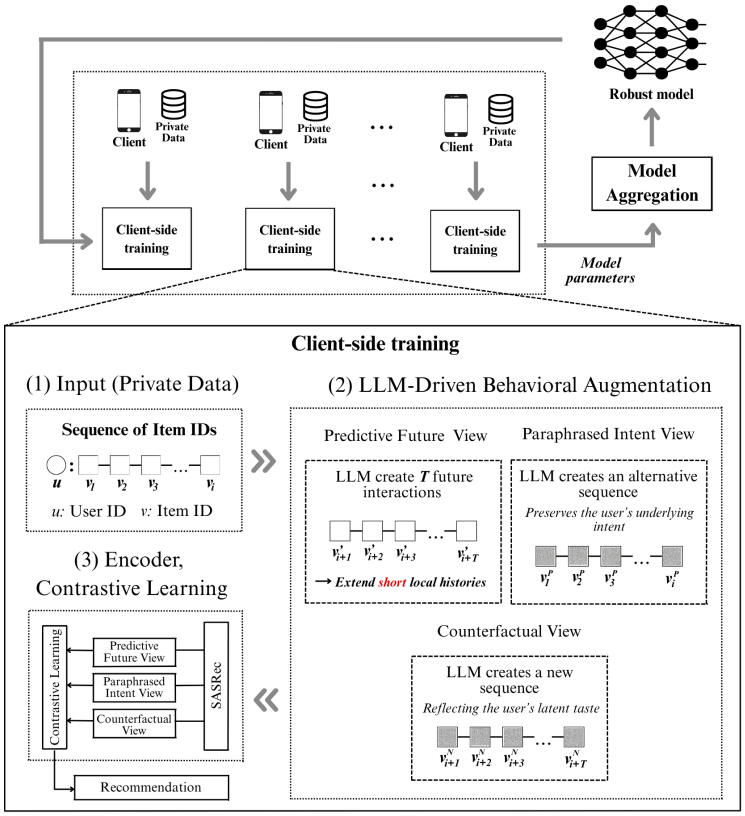


图 1. 所提出的 LUMOS 框架的整体架构。

释义意图保持视图。 该视图捕捉表达相同潜在偏好的其他方式，这些表达虽有所不同，但语义等价。给定原始序列 S_u，客户端会要求大语言模型在保留用户意图的前提下改写该序列，例如用相近的替代物品替换原有物品，或进行轻微调整顺序，从而生成连贯的释义序列。

$$S_u^{\mathrm{P}} = [\tilde{v}_{u,1}, \tilde{v}_{u,2}, \dots].$$

反事实负视图。反事实视图合成与用户偏好有意不一致的行为。在给定 S_u的条件下，客户端要求大语言模型生成一个对该用户而言不太可能出现的序列（例如来自相反类别或不相关领域），如下所示：

$$S_u^{\mathrm{N}} = [v_{u,1}^{-}, v_{u,2}^{-}, \dots].$$

对所有视图进行编码。所有序列均由同一个骨干网络编码：

$$h_u = f_\theta(S_u), \quad h_u^{\mathrm{F}} = f_\theta(S_u^{\mathrm{F}}), \quad h_u^{\mathrm{P}} = f_\theta(S_u^{\mathrm{P}}), \quad h_u^{\mathrm{N}} = f_\theta(S_u^{\mathrm{N}}).$$

这些表示构成了我们三视图对比目标的基础。

4.3 三视图对比学习

对比学习已被证明在集中式推荐场景中有效 [78,80,16]。在 LUMOS 中，我们通过构建利用大语言模型生成的视图的三视图目标，将这一思路应用于联邦序列场景。

对于每个用户 u，我们将真实序列 h_u 视为锚点，并定义两个正视图（未来视图和释义视图）以及一个负视图（反事实视图）：

$$\mathcal{P}_u = \{(h_u, h_u^{\mathrm{F}}), (h_u, h_u^{\mathrm{P}})\}, \qquad \mathcal{N}_u = \{(h_u, h_u^{\mathrm{N}})\}.$$

我们将正、负配分函数定义为：

$$\mathrm{pos}(u) = \sum_{(h_a, h_p) \in \mathcal{P}_u} \exp(\mathrm{sim}(h_a, h_p)/\tau), \tag{1}$$

$$\mathrm{neg}(u) = \sum_{(h_a, h_n) \in \mathcal{N}_u} \exp(\mathrm{sim}(h_a, h_n)/\tau). \tag{2}$$

因此，三视图对比损失为：

$$\mathcal{L}_u^{\mathrm{CL}} = -\log \frac{\mathrm{pos}(u)}{\mathrm{pos}(u) + \mathrm{neg}(u)}. \tag{3}$$

这里，$\mathrm{sim}(\cdot,\ \cdot)$ 是相似度函数（例如余弦相似度），τ 是温度超参数。直观而言，该损失鼓励 h_u 接近未来视图和释义视图，同时远离反事实视图。由于所有视图都是由大语言模型基于 S_u 条件生成的，因此得到的正、负视图在语义上信息丰富且具有用户特定性，当 S_u 较短时尤为有用。

4.4 本地目标与联邦训练

在每个参与的客户端上，总体训练目标将监督预测与三视图对比学习结合起来：

$$\mathcal{L}_u = \mathcal{L}_u^{\mathrm{Rec}} + \lambda_{\mathrm{CL}}\, \mathcal{L}_u^{\mathrm{CL}},$$

其中，λ_{CL} 控制对比项的强度。每个客户端使用全局参数 $\theta^{(t)}$初始化本地模型，执行若干轮随机梯度下降训练，以最小化 $\mathcal{L}_u$，然后上传更新后的参数 $\theta_u^{(t)}$至服务器。随后，服务器通过取平均值计算新的全局模型：

$$\theta^{(t+1)} = \frac{1}{|\mathcal{U}_t|} \sum_{u \in \mathcal{U}_t} \theta_u^{(t)}.$$

在 LUMOS 中，所有大语言模型调用、合成序列和三视图表示都仅限于客户端。服务器无需了解数据增强的具体细节，只负责协调标准的联邦优化；客户端目标则利用大语言模型生成的语义视图，从稀疏的交互历史中学习更丰富的序列表示。

5 实验

我们在三个真实世界的基准测试上对 LUMOS 进行了实证评估，并探讨以下问题：

RQ1： LUMOS 与最先进的推荐方法相比表现如何？**RQ2：** 大语言模型驱动的视图对性能有多大贡献？**RQ3：** LUMOS 对关键超参数有多敏感？

5.1 实验设置

数据集。 我们使用三个标准序列推荐数据集：亚马逊手机数据集、亚马逊婴儿用品数据集和 MIND [48,68], ，涵盖电子产品、婴儿用品和新闻。按照常见做法 [15,17], ，我们移除交互次数少于五次的用户和物品，按时间戳对每位用户的交互排序，将最后两次交互分别留作验证集和测试集，并将最大序列长度限制为 50（丢弃长度小于 5 的序列）。在联邦设置中，每位剩余用户对应一个客户端。

评估协议。 我们采用排名前 20 的命中率（HR@20）和排名前 20 的归一化折损累计增益（NDCG@20） [15,17]。HR@20 衡量真实的下一件物品是否出现在前 20 个结果中，而 NDCG@20 还会考虑该物品的排名位置。对于每位测试用户，我们对其此前未交互过的所有候选物品进行排序，并基于完整排序计算指标，以避免抽样偏差 [18]。
基线方法。我们将 LUMOS 与集中式方法和 FedSeqRec 进行比较：

– **集中式序列推荐系统**。 GRU4Rec [15] 是经典的基于GRU的会话模型，SASRec [17] 是单向自注意力序列推荐模型– **标准联邦序列推荐（FedSRS）**。FedSeq(GRU4Rec) 和 Fed- Seq(SASRec) 分别是 GRU4Rec 和 SASRec 的基础 FedAvg 版本：每个客户端在本地训练其骨干网络，服务器聚合参数，且不采用对比学习或大语言模型增强。– **对比式联邦序列推荐（ConFedSRS）**。ConFedSRS(GRU4Rec) 和 ConFedSRS(SASRec) 在 FedSRS 的基础上，为本地增强的序列（例如，经过掩码和裁剪的序列）加入标准的序列级对比损失，但不使用大语言模型。– **LUMOS（我们的方法）**。LUMOS 是一种联邦模型，采用 SASRec 骨干网络，且该网络始终保留在客户端。每个客户端使用本地（或具备隐私保护能力的）大语言模型，为自身序列生成三个行为视图（未来、释义改写、反事实），并在进行下一物品预测的同时优化三视图对比损失。

实现细节。 所有模型均使用 PyTorch 实现，并在配备 NVIDIA A100 GPU 的服务器上训练。除非另有说明，LUMOS 使用 SASRec 作为其序列骨干网络。我们采用同步联邦平均，每个通信轮次采样 10% 的客户端；每个被选中的客户端运行五个本地

表1. 三个数据集上的Top-K 推荐性能。每列中最佳的集中式结果以下划线标出，最佳的联邦式结果以**粗体**显示。

方法	手机		Baby		MIND	
	HR@20	NDCG@20	HR@20	NDCG@20	HR@20	NDCG@20
GRU4Rec	0.0700	0.0290	0.0370	0.0140	0.1310	0.0580
SASRec	0.0840	0.0370	0.0390	0.0150	0.1510	0.0690
FedSeq(GRU4Rec)	0.0500	0.0180	0.0230	0.0088	0.0700	0.0240
FedSeq(SASRec)	0.0510	0.0185	0.0290	0.0102	0.0650	0.0228
ConFedSRS(GRU4Rec)	0.0860	0.0385	0.0340	0.0142	0.1300	0.0610
ConFedSRS(SASRec)	0.0950	0.0425	0.0420	0.0173	0.1370	0.0620
LUMOS	**0.1570**	**0.0670**	**0.0680**	**0.0250**	**0.1660**	**0.0730**

训练轮次后再进行聚合，我们共训练 100 轮。我们使用 Adam，学习率为 1×10^{-3}，权重衰减为 1×10^{-5}，批量大小为 128，并进行最大范数为 5.0 的梯度裁剪。对比损失权重设为 $\lambda_{\mathrm{CL}} = 0.1$，温度参数设为 $\tau = 0.07$。

5.2 总体推荐性能（RQ1）

表1报告了所有数据集上的 HR@20 和 NDCG@20。上方部分列出了中心化模型，中间部分包含标准 FedSeq 基线，下方部分则展示了 ConFedSRS 和我们的 LUMOS 框架。每列中最佳的集中式结果以带下划线标示，最佳的联邦式结果以**粗体**标示。

在联邦设置下，FedSeq 的两个变体都因数据划分以及缺少辅助自监督而导致 HR@20 和 NDCG@20 明显下降。加入简单的序列级对比目标（ConFedSRS）后，性能始终优于 FedSeq，这表明即便是人工设计的数据增强也有助于稳定客户端表示。LUMOS 在三个数据集和各项指标上均取得最佳联邦性能，且显著优于 ConFedSRS。与最强基线 ConFedSRS(SASRec) 相比，LUMOS 的 HR@20 提高了 6.2–7.6 个百分点，NDCG@20 提高了约 1.0 个百分点，这得益于大语言模型生成的未来行为、释义行为和反事实行为所提供的丰富三视图监督。值得注意的是，LUMOS 在所有数据集上的表现也超过了中心化 SASRec，这表明在联邦训练中，基于语义的大语言模型视图可以弥补无法直接访问全局交互的不足。

表2。 LUMOS 在三个数据集上的消融研究。最佳结果以粗体标出。

变体	手机		Baby		MIND	
	HR@20	NDCG@20	HR@20	NDCG@20	HR@20	NDCG@20
LUMOS（完整）	**0.1550**	**0.0660**	**0.0650**	**0.0240**	**0.1620**	**0.0725**
不含预测的未来视图	0.1410	0.0590	0.0590	0.0215	0.1490	0.0670
不使用释义意图视图	0.1360	0.0565	0.0570	0.0205	0.1450	0.0655
不使用反事实视图	0.1320	0.0550	0.0555	0.0200	0.1425	0.0648

5.3 消融研究（RQ2）

表2显示，移除 LUMOS 的任何一个组件都会导致性能明显下降。去掉预测性未来视图造成的降幅最大

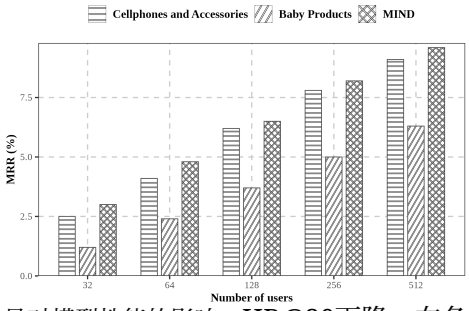


图2. 参与用户数量对模型性能的影响。HR@20下降，在各数据集上降幅约为0.01–0.015。移除释义视图或反事实视图也会造成类似下降（约为0.008–0.012），这表明二者提供了互补的语义信号。去除时间正则项后，HR@20还会进一步下降约0.005–0.008，证实了该正则项在稳定用户表示方面的作用。总体而言，完整的三视图设计和时间一致性是实现最佳准确率的必要条件。

5.4 超参数分析（RQ3）

每个 epoch 的用户数量的影响。如图2所示，随着参与的客户端增多，LUMOS 的性能持续提升。MRR（平均倒数排名）用于衡量正确的下一项在每个用户的候选项中的排名高低；在 32 个用户时，MRR 较低（约为 2–3%），但随着用户数增至 64 和 128，MRR 稳步上升；在 512 个用户时，CellPhone 上超过 8%，MIND 上则接近 9%。联邦训练的表现也越来越优于单独训练，在最大规模下，MRR 差距超过 3 个百分点。这证实了更广泛的参与能提供更丰富的行为信号，从而增强序列建模。

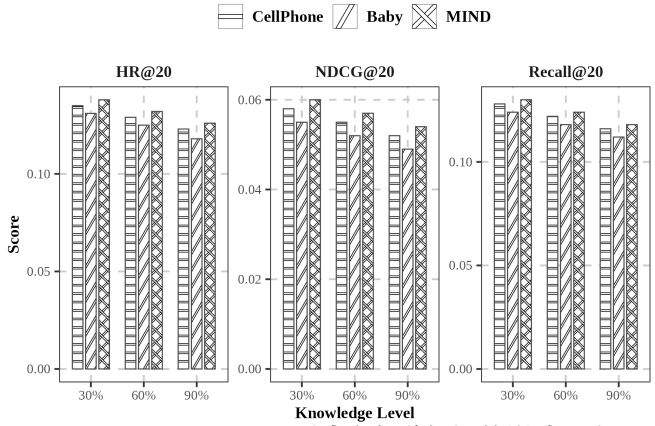


图3. 部分知识对模型效用的影响。**噪声和部分知识的影响**。如图3所示，即使对抗性知识从30%增加到90%，LUMOS仍保持稳定。在所有数据集上，HR@20仅下降约0.01–0.015，而NDCG@20和Recall@20的降幅不到0.01。这些微小变化表明，语义——

基于大语言模型的三视图增强所引入的结构有效限制了知情对抗噪声的影响 [50]。

6 结论

我们从表征学习的角度重新审视了联邦序列推荐，以应对数据稀疏性和增强方式有限的问题。我们提出了 LUMOS，利用端侧大语言模型作为语义行为生成器来增强本地学习。LUMOS采用三视图对比目标，结合本地生成的预测视图、释义视图和反事实视图。这实现了多正样本InfoNCE损失，在保护隐私的同时对齐语义行为，避免向服务器发生任何数据泄露。未来的工作可以探索个性化提示或适配器微调 [49,4,57,43],，以及纳入更丰富的信号 [56,28,41,37,39]。此外，整合通信与能源感知的成本模型，将有助于在资源受限设备上进行部署，推动隐私保护的大语言模型驱动的联邦推荐 [14,38,40,13,46]。

参考文献

1. Ammad-Ud-Din, M., Ivannikova, E., Khan, S.A., Oyomno, W., Fu, Q., Tan, K.E., Flanagan, A.: 面向隐私保护个性化推荐系统的联邦协同过滤。arXiv 预印本 arXiv:1901.09888（2019） 2. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT：用于语言理解的深度双向 Transformer 预训练。载于：NAACL-HLT。第 4171–4186 页（2019） 3. Duong, C.T., Nguyen, T.T., Hoang, T.D., Yin, H., Weidlich, M., Nguyen, Q.V.H.: Deep mincut：通过检测社区学习节点嵌入。Pattern Recognition，第 109126 页（2022） 4. Duong, C.T., Nguyen, T.T., Hoang, T.D., Yin, H., Weidlich, M., Nguyen, Q.V.H.: Deep mincut：通过检测社区学习节点嵌入。Pattern Recognition **134**，109126（2023） 5. Duong, C.T., Nguyen, T.T., Yin, H., Weidlich, M., Mai, T.S., Aberer, K., Nguyen, Q.V.H.: 通过异构网络嵌入实现高效且有效的多模态查询。IEEE Transactions on Knowledge and Data Engineering **34**(11), 5307–5320（2022） 6. Fang, H., Zhang, D., Shu, Y., Guo, G.: 序列推荐的深度学习：算法、影响因素与评估。TOIS **39**(1)，1–42（2020） 7. He, X., Deng, K., Wang, X., Li, Y., Zhang, Y., Wang, M.: LightGCN：简化并增强用于推荐的图卷积网络。载于：SIGIR。第 639–648 页（2020） 8. He, X., Liao, L., Zhang, H., Nie, L., Hu, X., Chua, T.S.: 神经协同过滤。载于：WWW。第 173–182 页（2017） 9. Hung, N.Q.V., Thang, D.C., Tam, N.T., Weidlich, M., Aberer, K., Yin, H., Zhou, X.: 以最少投入验证通用众包任务的答案。The VLDB Journal **26**，855–880（2017） 10. Hung, N.Q.V., Weidlich, M., Tam, N.T., Miklós, Z., Aberer, K., Gal, A., Stantic, B.: 在数据模型的按需协调中处理概率完整性约束。Information Systems**83**，166–180（2019）

11. Huynh, T.T., Duong, C.T., Nguyen, T.T., Van, V.T., Sattar, A., Yin, H., Nguyen, Q.V.H.: 基于整体嵌入的网络对齐。TKDE **35**(2)，1881–1894（2021） 12. Huynh, T.T., Nguyen, M.H., Nguyen, T.T., Nguyen, P.L., Weidlich, M., Nguyen, Q.V.H., Aberer, K.: 在股票走势预测中高效整合多阶动态与内部动态。载于：第十六届 ACM 网络搜索与数据挖掘国际会议论文集。第 850–858 页（2023） 13. Huynh, T.T., Nguyen, T.B., Nguyen, P.L., Nguyen, T.T., Weidlich, M., Nguyen, Q.V.H., Aberer, K.: Fast-fedul：一种无需训练且具备可证明偏斜鲁棒性的联邦遗忘方法。载于：ECML-PKDD。第 55–72 页（2024） 14. Huynh, T.T., Nguyen, T.B., Nguyen, T.T., Nguyen, P.L., Yin, H., Nguyen, Q.V.H., Nguyen, T.T.: 联邦推荐的可认证遗忘。TOIS **43**(2)，1–29（2025） 15. Jannach, D., Ludewig, M.: 当循环神经网络遇上邻域：会话推荐。载于：RecSys。第 306–310 页（2017） 16. Jing, M., Zhu, Y., Zang, T., Wang, K.: 推荐系统中的对比式自监督学习：综述。TOIS **42**(2)，1–39（2023） 17. Kang, W.C., McAuley, J.: 自注意力序列推荐。载于：ICDM。第 197–206 页（2018） 18. Krichene, W., Rendle, S.: 关于物品推荐的采样指标。载于：KDD。第 1748–1757 页（2020） 19. Lin, Z., Pan, W., Yang, Q., Ming, Z.: 一种通过伪标记和秘密共享实现的通用联邦推荐框架。TOIS **41**(2)，1–37（2022） 20. Liu, Y., Tan, B., Zheng, V.W., Chen, K., Yang, Q.: 联邦推荐系统。载于：《联邦学习：隐私与激励》，第 225–239 页（2020） 21. Long, J., Chen, T., Ye, G., Zheng, K., Nguyen, Q.V.H., Yin, H.: 去中心化兴趣点推荐中的物理轨迹推断攻击与防御。载于：WWW。第 3379–3387 页（2024） 22. Long, J., Ye, G., Chen, T., 等：基于扩散的云边端协同学习，用于下一兴趣点推荐。载于：KDD。第 2026–2036 页（2024） 23. Luo, L., Liu, B.: 用于联邦社交推荐的双重对比方法。载于：IJCNN。第 1–8 页（2022） 24. Luo, S., Xiao, Y., Zhang, X., Liu, Y., Ding, W., Song, L.: Perfedrec++：通过自监督预训练增强个性化联邦推荐。TIST**15**(5)，1–24（2024） 25. Lyu, H., Jiang, S., Zeng, H., Xia, Y., Wang, Q., Zhang, S., Chen, R., Leung, C., Tang, J., Luo, J.: LLM-rec：通过提示大语言模型实现个性化推荐。载于：NAACL。第 583–612 页（2024） 26. Min, B., Ross, H., Sulem, E., Veyseh, A.P.B., Nguyen, T.H., Sainz, O., Agirre, E., Heintz, I., Roth, D.: 基于大型预训练语言模型的自然语言处理最新进展：综述。CSUR **56**(2)，1–40（2023） 27. Minto, L., Haller, M., Livshits, B., 等：通过隐式反馈增强联邦协同过滤的隐私保护。载于：RecSys。第 342–350 页（2021） 28. Nguyen, D.D.A., Nguyen, M.H., Nguyen, P.L., Jo, J., Yin, H., Nguyen, T.T.: 基于位置的社交网络中异构超图表示的多任务学习。载于：ADMA。第 161–177 页（2024） 29. Nguyen, M.H., Huynh, T.T., Nguyen, T.T., Nguyen, T., Nguyen, P.L., Pham, H.T., Jo, J.: 基于异构联邦块网络的端侧诊断推荐。SCIS **68**(4)，140102（2025）

30. Nguyen, Q.V.H., Do, S.T., Nguyen, T.T., Aberer, K.：基于标签的论文检索：通过多样性感知减少用户工作量。载于：高级应用数据库系统国际会议。第510–528页（2015） 31. Nguyen, Q.V.H., Duong, C.T., Nguyen, T.T., Weidlich, M., Aberer, K., Yin, H., Zhou, X.：通过众包发现论点。VLDBJ**26**，511–535（2017） 32. Nguyen, Q.V.H., Duong, C.T., Nguyen, T.T., Weidlich, M., Aberer, K., Yin, H., Zhou, X.：通过众包发现论点。The VLDB Journal **26**(4)，511–535（2017） 33. Nguyen, Q.V.H., Nguyen, T.T., Chau, V.T., Wijaya, T.K., Miklós, Z., Aberer, K., Gal, A., Weidlich, M.：Smart：用于分析和协调模式匹配网络的工具。载于：ICDE。第1488–1491页（2015） 34. Nguyen, Q.V.H., Nguyen Thanh, T., Miklós, Z., Aberer, K.：通过众包协调模式匹配网络。EAI Endorsed Transactions on Collab- orative Computing **1**(2)，e2（2014） 35. Nguyen, Q.V.H., Zheng, K., Weidlich, M., Zheng, B., Yin, H., Nguyen, T.T., Stantic, B.：具有冲突目标的假设分析：为探索推荐数据范围。载于：2018年IEEE第34届数据工程国际会议（ICDE）。第89–100页。IEEE（2018） 36. Nguyen, T.T., Nguyen, Q.V.H., Nguyen, T.T.等：操纵推荐系统：投毒攻击及对策综述。CSUR **57**，1–39（2025） 37. Nguyen, T.T., Huynh, T.T., Pham, M.T., Hoang, T.D., Nguyen, T.T., Nguyen, Q.V.H.：利用混合生成对抗网络验证功能冗余。KBS **264**，110342（2023） 38. Nguyen, T.T., Huynh, T.T., Ren, Z., Nguyen, T.T., Nguyen, P.L., Yin, H., Nguyen, Q.V.H.：隐私保护的可解释人工智能：综述。SCIS **68**(1)，111101（2025） 39. Nguyen, T.T., Huynh, T.T., Yin, H., Weidlich, M., Nguyen, T.T., Mai, T.S., Nguyen, Q.V.H.：利用大规模流数据实现具有延迟保证的谣言检测。VLDBJ**32**(2)，369–387（2023） 40. Nguyen, T.T., Nguyen, T.T., Weidlich, M., Jo, J., Nguyen, Q.V.H., Yin, H., Liew, A.W.C.：利用分区图变换器处理推荐系统中的低同质性。TKDE（2024） 41. Nguyen, T.T., Phan, T.C., Nguyen, M.H., Weidlich, M., Yin, H., Jo, J., Nguyen, Q.V.H.：面向流式谣言图的模型无关多样化解释。KBS **253**，109438（2022） 42. Nguyen, T.T., Phan, T.C., Pham, H.T., Nguyen, T.T., Jo, J., Nguyen, Q.V.H.：基于示例的图上流式欺诈检测解释。Information Sciences **621**，319–340（2023） 43. Nguyen, T.T., Ren, Z., Nguyen, T.T., Jo, J., Nguyen, Q.V.H., Yin, H.：针对多模态异质性的可移植基于图的谣言检测。KBS **284**，111310（2024） 44. Nguyen, T.T., Ren, Z., Pham, T., Nguyen, P.L., Nguyen, Q.V.H., Yin, H.：指令引导的图像编辑综述。EAAI（2026） 45. Nguyen, T.T., Weidlich, M., Yin, H., Zheng, B., Nguyen, Q.H., Nguyen, Q.V.H.：Factcatch：以最少用户工作量实现增量式按需事实核查。载于：第43届国际ACM SIGIR信息检索研究与发展会议论文集。第2165–2168页（2020） 46. Nguyen, T.T., Nguyen, T.T., Nguyen, T.H., Yin, H., Nguyen, T.T., Jo, J., Nguyen, Q.V.H.：用于渐进式最大频繁子图挖掘的同构图嵌入。TIST **15**(1)，1–26（2023）

47. Nguyen Thanh, T., Quach, N.D.K., Nguyen, T.T., Huynh, T.T., Vu, V.H., Nguyen, P.L., Jo, J., Nguyen, Q.V.H.：利用基于生成式代理的攻击对基于图神经网络的推荐系统实施投毒攻击。ACM信息系统汇刊**41**(3)，1–24（2023） 48. Ni, J., Li, J., McAuley, J.：利用远程标注评论和细粒度方面为推荐提供依据。载于：EMNLP-IJCNLP。第188–197页（2019） 49. Pham, K.T., Nguyen, T.H., Jo, J., Nguyen, Q.V.H., Nguyen, T.T.：多语言文本到SQL：通过基准测试探究语言模型与协作式语言智能体的能力边界。载于：ADC。第108–123页（2025） 50. Pham, M.T., Huynh, T.T., Nguyen, T.T., Nguyen, T.T., Jo, J., Yin, H., Nguyen, Q.V.H.：人脸伪造与人脸取证双重基准测试研究。CAAI-TIT **9**(6)，1377–1397（2024） 51. Qu, L., Tang, N., Zheng, R., Nguyen, Q.V.H., Huang, Z., Shi, Y., Yin, H.：用于推荐的半去中心化联邦自我图学习。载于：WWW。第339–348页（2023） 52. Qu, L., Yuan, W., Zheng, R., Cui, L., Shi, Y., Yin, H.：迈向个性化隐私：用户主导的联邦推荐数据贡献。载于：WWW。第3910–3918页（2024） 53. Ren, Z., Nguyen, T.T., Nejdl, W.：用于可解释呼吸音分析的原型学习。载于：ICASSP论文集。第9087–9091页（2022） 54. Rendle, S., Freudenthaler, C., Schmidt-Thieme, L.：为下一购物篮推荐分解个性化马尔可夫链。载于：WWW。第811–820页（2010） 55. Rong, D., He, Q., Chen, J.：联邦学习场景下对基于深度学习的推荐模型实施投毒攻击。载于：IJCAI。第2204–2210页（2022） 56. Sakong, D., Nguyen, T.T.：利用小波超图扩散处理推荐系统中的异配性。arXiv预印本 arXiv:2501.14399（2025） 57. Sakong, D., Vu, V.H., Huynh, T.T., Le Nguyen, P., Yin, H., Nguyen, Q.V.H., Nguyen, T.T.：利用异构超图多重注意力进行高阶知识增强推荐。Information Sciences **680**，121165（2024） 58. Shehmir, S., Kashef, R.：LLM4Rec：大语言模型在推荐系统中集成应用的全面综述——方法、应用与挑战。Future Internet **17**(6)（2025） 59. Sun, F., Liu, J., Wu, J., Pei, C., Lin, X., Ou, W., Jiang, P.：Bert4rec：基于Transformer双向编码器表示的序列推荐。载于：CIKM。第1441–1450页（2019） 60. Tang, Z., Huan, Z., Li, Z., Zhang, X., Hu, J., Fu, C., Zhou, J., Zou, L., Li, C.：一个模型适用于所有场景：大语言模型是领域无关推荐系统。TOIS **43**(5)，1–27（2025） 61. Thang, D.C., Dat, H.T., Tam, N.T., Jo, J., Hung, N.Q.V., Aberer, K.：先天还是后天：图神经网络的特征与结构。PRL **159**，46–53（2022） 62. Thang, D.C., Tam, N.T., Hung, N.Q.V., Aberer, K.：多样化技术评估。载于：数据库与专家系统应用国际会议。第215–231页（2015） 63. Toan, N.T., Cong, P.T., Tam, N.T., Hung, N.Q.V., Stantic, B.：多样化群组推荐。IEEE Access **6**，17776–17786（2018） 64. Trung, H.T., Van Vinh, T., Tam, N.T., Jo, J., Yin, H., Hung, N.Q.V.：在基于位置的社交网络中，通过高阶、动态和多角色上下文学习整体交互。IEEE Transactions on Knowledge and Data Engineering **35**(5)，5002–5016（2022） 65. Wang, H., Xu, Y., Yang, C., Shi, C., Li, X., Guo, N., Liu, Z.：面向推荐的知识自适应对比学习。载于：WSDM。第535–543页（2023）

66. Wang, Z., Yu, J., Gao, M., 等：推荐系统中的投毒攻击与防御：综述。arXiv 预印本 arXiv:2406.01022（2024）67. Wu, C., Wu, F., Qi, T., 等：Fedcl：用于隐私保护推荐的联邦对比学习。arXiv 预印本 arXiv:2204.09850（2022）68. Wu, F., Qiao, Y., Chen, J.H., Wu, C., Qi, T., Lian, J., Liu, D., Xie, X., Gao, J., Wu, W., 等：Mind：用于新闻推荐的大规模数据集。载于：ACL。第 3597–3606 页（2020）69. Wu, J., Wang, X., Feng, F., 等：用于推荐的自监督图学习。载于：SIGIR。第 726–735 页（2021）70. Wu, L., Zheng, Z., Qiu, Z., Wang, H., Gu, H., Shen, T., Qin, C., Zhu, C., Zhu, H., Liu, Q., Xiong, H., Chen, E.：面向推荐的大语言模型综述。WWW **27**(5), 60（2024）71. Wu, Y., Xie, R., Zhu, Y., Ao, X., Chen, X., Zhang, X., Zhuang, F., Lin, L., He, Q.：推荐中的多视图多行为对比学习。载于：DASFAA。第 166–182 页（2022）72. Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., Yu, P.S.：图神经网络全面综述。TNNLS **32**(1), 4–24（2020）73. Xia, X., Yin, H., Yu, J., 等：用于会话推荐的自监督超图卷积网络。载于：AAAI。第 4503–4511 页（2021）74. Xie, X., Sun, F., Liu, Z., Wu, S., Gao, J., Zhang, J., Ding, B., Cui, B.：序列推荐的对比学习。载于：ICDE。第 1259–1273 页（2022）75. Yao, Y., Zhang, J., Wu, J., Huang, C., Xia, Y., Yu, T., Zhang, R., Kim, S., Rossi, R., Li, A., Yao, L., McAuley, J., Chen, Y., Joe-Wong, C.：联邦大语言模型：当前进展与未来方向（2024）76. Ye, R., Ge, R., Zhu, X., Chai, J., Du, Y., Liu, Y., Wang, Y., Chen, S.：FedLLM-基准：大语言模型联邦学习的现实基准测试（2024）77. Yin, H., Qu, L., Chen, T., Yuan, W., Zheng, R., Long, J., Xia, X., Shi, Y., Zhang, C.：端侧推荐系统：全面综述。arXiv 预印本 arXiv:2401.11441（2024）78. Yu, J., Xia, X., Chen, T., Cui, L., Nguyen, Q.V.H., Yin, H.：Xsimgcl：迈向极简的推荐图对比学习。TKDE（2023）79. Yu, J., Yin, H., Li, J., 等：用于社交推荐的自监督多通道超图卷积网络。载于：WWW。第 413–424 页（2021）80. Yu, J., Yin, H., Xia, X., Chen, T., Li, J., Huang, Z.：推荐系统自监督学习综述。TKDE（2023）81. Yu, J., Yin, H., Xia, X., 等：图增强是否必要？用于推荐的简单图对比学习。载于：SIGIR。第 1294–1303 页（2022）82. Yu, Y., Liu, Q., Wu, L., 等：通过投毒物品嵌入攻击联邦推荐系统的无目标攻击及其防御。载于：AAAI。第 4854–4863 页（2023）83. Yuan, W., Nguyen, Q.V.H., He, T., 等：操纵联邦推荐系统：利用合成用户进行投毒及其对策。载于：SIGIR。第 1690–1699 页（2023）84. Yuan, W., Qu, L., Cui, L., Tong, Y., Zhou, X., Yin, H.：Hetefedrec：具有模型异质性的联邦推荐系统。arXiv 预印本 arXiv:2307.12810（2023）85. Yuan, W., Yang, C., Nguyen, Q.V.H., Cui, L., He, T., Yin, H.：针对联邦推荐系统的交互级成员推理攻击。载于：WWW。第 1053–1062 页（2023）

86. Yuan, W., Yang, C., Qu, L., Hung Nguyen, Q.V., Ye, G., Yin, H.：Ptf-fsr：无参数传输联邦序列推荐系统。TOIS **43**(2)（2025）87. Yuan, W., Yang, C., Qu, L., 等：隐藏你的模型：无参数传输联邦推荐系统。arXiv 预印本 arXiv:2311.14968（2023）88. Yuan, Z., Yuan, F., Song, Y., Li, Y., Fu, J., Yang, F., Pan, Y., Ni, Y.：推荐系统的下一步发展方向是什么？重新审视基于 ID 与基于模态的推荐模型。载于：SIGIR。第 2639–2649 页（2023）89. Zhang, S., Yuan, W., Yin, H.：针对属性推断攻击的联邦推荐系统全面隐私分析。TKDE（2023）90. Zhao, B., van der Aa, H., Nguyen, T.T., Nguyen, Q.V.H., Weidlich, M.：Eires：事件流处理中远程数据的高效集成。载于：2021年国际数据管理会议论文集。第 2128–2141 页（2021）91. Zhao, Z., Fan, W., Li, J., Liu, Y., Mei, X., Wang, Y., Wen, Z., Wang, F., Zhao, X., Tang, J., 等：大语言模型（llms）时代的推荐系统。TKDE（2024）