MergeRec：面向数据隔离跨域序列推荐的模型合并

Hyunsoo Kim*成均馆大学 水原，大韩民国
khs1778@skku.edu

Jaewan Moon*成均馆大学 水原，大韩民国
jaewan7599@skku.edu

Seongmin Park 成均馆大学 水原，大韩民国
psm1206@skku.edu

Jongwuk Lee†成均馆大学 水原，大韩民国
jongwuklee@skku.edu

arXiv:2601.01753v1 [cs.IR] 5 Jan 2026

摘要

在特定领域数据上训练的现代推荐系统通常难以在多个领域之间实现泛化。跨域序列推荐已成为应对这一挑战的一个有前景的研究方向；然而，现有方法面临一些根本性限制，例如依赖跨领域的重叠用户或物品，或作出忽视隐私约束的不现实假设。在本文中，我们提出了一个新的框架 ***MergeRec***，该框架基于模型合并，针对一种新的、更符合现实的问题设定，即数据隔离的跨域序列推荐：在该设定下，原始用户交互数据无法在不同领域之间共享。MergeRec 由三个关键组件构成：(1)合并初始化，(2) 伪用户数据构建，以及 (3) 协同合并优化。首先，我们使用免训练合并技术初始化合并模型。接下来，我们将每个物品视为各领域中的一个虚拟序列，从而构建伪用户数据，在不依赖真实用户交互的情况下生成有意义的训练样本。最后，我们通过联合目标优化特定领域合并权重。该联合目标结合了推荐损失与蒸馏损失：前者促使合并模型识别相关物品，后者则将协同过滤信号从微调后的源模型迁移过来。大量实验表明，MergeRec 不仅保留了原始模型的优势，还显著提升了其在未见领域上的泛化能力。与传统的模型合并方法相比，MergeRec 始终取得更优的性能，在 Recall@10 上平均提升最高达 17.21%，凸显了模型合并作为一种可扩展且有效的方法，在构建通用推荐系统方面的潜力。源代码可在 github.com/DIALLab-SKKU/MergeRec 获取。

CCS 概念

• **信息系统** →**推荐系统**。

关键词

跨域序列推荐；模型合并；数据隔离；任务向量

*两位作者对本研究贡献相同。
†通讯作者

本作品依据知识共享署名 4.0 国际许可协议授权。*KDD 2026*，大韩民国济州岛。
© 2026 版权所有归所有者/作者所有。ACM ISBN
979-8-4007-2258-5/2026/08 https://doi.org/10.1145/3770854.3780264

ACM 引用格式：
Hyunsoo Kim、Jaewan Moon、Seongmin Park 和 Jongwuk Lee。2026。MergeRec：面向数据隔离跨域序列推荐的模型合并。载于《第*32*届 *ACM SIGKDD* 知识发现与数据挖掘大会论文集》第*1*卷（*KDD 2026*），*2026*年*8*月*9*日至*13*日，济州岛，大韩民国。ACM，纽约，纽约州，美国，14页。https://doi.org/10.1145/3770854.3780264

1 引言

序列推荐（SR）旨在根据用户的交互历史，预测用户可能偏好的下一批物品。近期的神经序列推荐模型 [6, 12, 14, 15, 33] 采用了多种架构，以有效捕捉物品之间的序列依赖关系。然而，它们仍面临冷启动和数据稀疏问题等固有挑战 [22–24, 29], ，这些问题限制了其泛化能力和总体性能。

跨域序列推荐（CDSR）已成为一个颇具前景的研究方向 [3]。CDSR 旨在通过在多个领域间联合训练模型 [26, 27, 37, 52]，或将知识从数据丰富的领域迁移到数据较稀疏的领域 [1, 16, 17, 20]，来提高推荐准确率。然而，现有的 CDSR 研究面临三项根本性局限。（1）**用户/物品重叠依赖**：知识迁移通常依赖于不同领域之间存在重叠用户或物品。然而，实际中的此类重叠极其有限。我们发现，在八个亚马逊领域中，仅有 16 个用户和 0 个物品是共享的，这反映了各领域独立运营的现实情况；（2）**数据隔离**：在现实场景中，由于组织边界或隐私法规 [41, 47, 48]，原始用户数据的访问往往受到限制。用户日志包含敏感信息，受隐私限制，无法在不同领域之间共享，导致领域专属训练数据无法获取；（3）**可扩展性低**：在多个领域间进行联合训练会产生大量计算开销，使其难以用于大规模部署。因此，大多数以往研究仅限于整合两个或三个领域，尚未解决可扩展的多领域整合问题。

我们提出，模型合并 [4, 5, 8–11、18、25、31、32、38、40、43–46, 49, 50] 为构建通用推荐系统提供了一种有效的替代方案。模型合并将多个领域专属模型或任务专属模型中的微调参数整合到单一统一模型中。这一范式具有多项优势，能够直接应对跨域序列推荐的关键局限：（1）无需不同领域之间存在重叠用户或物品；（2）能够自然地保护用户隐私，因为只需使用模型参数，而不需要敏感交互数据；（3）通过避免跨域联合训练所带来的计算负担，实现了高可扩展性。



图 1：多个领域中的序列推荐训练用户序列与测试用户序列示意图。每个测试用户序列（黑框）都包含此前的所有交互，包括训练期间的交互（蓝框），这表明在真实场景中，测试数据是训练数据的超集。

本文探讨了在一种称为数据隔离的跨域序列推荐的新型现实问题设定下应用模型合并的可行性。该设定源于实际世界中的约束：用户交互数据只能用于训练领域专属模型，不能在不同领域之间共享，也无法在训练后再次访问。不同于通常依赖强且不切实际假设（即，存在重叠用户或物品）的传统跨域序列推荐，数据隔离的跨域序列推荐允许不同领域彼此不相交。此外，隐私保护的跨域序列推荐通常要求在模型优化过程中访问领域专属的交互数据，而数据隔离的跨域序列推荐无需访问任何用户交互数据，即可构建通用的跨域推荐系统，从而为用户隐私提供更强的保障。

然而，在这种设置下，直接应用现有的模型合并方法并非易事，主要有两个原因。首先，由于不同领域之间不共享交互数据，因此无法采用模型合并范式中常用于优化合并权重的测试时自适应方案。其次，即使可以访问测试数据，在序列推荐中利用测试序列也会违背模型合并的核心假设。模型合并范式明确禁止使用训练数据，但这些假设并不适用于序列推荐系统。在这类系统中，测试序列并非独立于训练数据，而是由同一不断演变的用户行为生成的。因此，在合并过程中使用测试时的用户交互序列，必然会暴露训练信息（图 1）。因此，利用测试数据进行模型合并与数据隔离的跨域序列推荐设置根本不相容。

为此，我们提出了***MergeRec***，这是一个专为数据隔离的跨域序列推荐设计的新型框架。MergeRec 包含三个关键组成部分：(1) 合并初始化，(2) 伪用户数据构建，以及 (3) 协同合并优化。首先，我们采用基于任务向量的免训练合并方法合成一个初始合并模型。任务向量定义为微调模型与其对应的预训练模型之间的参数差异，用于捕获领域特定知识 [9]。接下来，我们将每个领域中的每个物品视为一个独立序列，以此构建伪用户数据。尽管方法简单，MergeRec 仍能在不依赖真实用户数据的情况下，为合并不同领域构建有意义的样本，从而有效模拟跨领域的冷启动用户。最后，我们通过面向推荐的合并目标优化特定领域合并权重。

为了设计一种用于合并推荐系统的有效目标函数，我们认为理想的合并模型应满足两项基本要求。首先，它应能够解码用户的多种意图，这些意图通常体现在领域特定的序列模式中。其次，统一模型应具备强大的排序能力，能够在每个领域情境中准确地将点击概率最高的物品排在前面。我们指出，现有的自适应合并方法，即*AdaMerging*，只关注后者，因此不足以用于合并推荐系统（第 3 节）。

为克服这一局限，我们提出了一个联合目标，其中结合了：(1) 蒸馏损失，它利用微调模型的预测分布作为软标签；以及 (2) 推荐损失，它将微调模型针对其对应领域中的伪用户预测的排名第一的物品作为硬标签。蒸馏损失将微调模型中的协同过滤（CF） [28] 信号迁移至合并模型，而推荐损失则引导合并模型根据物品被点击的可能性对其进行准确排序。

大量实验表明，MergeRec 不仅保留了各个源模型的优势，还能有效泛化到未见领域。与现有合并方法和强基线方法（包括微调模型和联合学习模型）相比，MergeRec 始终表现出色。具体而言，在 Recall@10 上，MergeRec 相较于联合学习和 AdaMerging 的平均提升分别为 8.72% 和 17.21%。这些结果表明，模型合并可以成为构建通用推荐系统的一种可扩展且高效的范式。

我们的贡献可概括如下：

- **全面的实证分析**：我们提供了实证证据，表明基于熵的优化虽然在计算机视觉和自然语言处理中有效，却无法从根本上捕捉推荐系统中固有的多意图行为模式。

- **首个面向推荐系统的模型合并框架**：我们提出了***MergeRec***，一种专为推荐系统设计的基于任务向量的模型合并框架。MergeRec 包含三个关键组件：（1）免训练合并初始化；（2）隐私保护的伪用户数据构造；以及（3）面向推荐的合并目标。
- **全面评估**：通过在八个亚马逊基准数据集和四种骨干架构上开展大量实验，我们证明 MergeRec 始终优于现有的模型合并基线方法。值得注意的是，MergeRec 对未见过的领域表现出更强的泛化能力，并且在数据稀缺条件下性能稳健。

2 预备知识

2.1 跨域序列推荐

令 $\mathbb{D} = \{D_1, D_2, \ldots\ldots, D_K\}$ 表示所有推荐域的集合，其中 D_k 表示第 k 个领域。每个领域 D_k 由一组物品 $\mathbb{I}_k$ 和一组用户 $\mathbb{U}_k$ 构成。对于任意用户 $u \in \mathbb{U}_k$，其交互历史表示为依据时间戳排列的物品序列：$u = \left[i_1, i_2, \ldots, i_{|u|}\right]$，其中 $|u|$ 表示用户 u 的交互次数。跨域序列推荐模型旨在通过估计用户 u 在给定其过往交互的条件下接下来与各物品交互的概率，预测并对 $\mathbb{I}_k$ 中的物品进行排序：

$$\theta^* = \arg\max_{\theta} P\left(i = i_{|u|+1} \mid u, \theta\right), \tag{1}$$

其中 θ 表示跨域序列推荐模型的参数。

2.2 基于文本的序列推荐

基于文本的序列推荐 [7, 13] 利用预训练语言模型（PLM）对物品级文本信息进行编码。通过使用文本描述表示用户和物品，这种方法能够为此前未见过的物品（即冷启动物品）提供推荐。

形式上，物品 t_i 的文本表示由其属性描述（例如标题、品牌和类别）构成。用户 t_u 的文本表示定义为该用户交互过的所有物品的文本表示的拼接：

$$t_u = \left[t_{i_1}; t_{i_2}; \ldots; t_{i_{|u|}}\right], \tag{2}$$

其中；表示拼接运算符 .

令 $f(\cdot \mid \theta_k)$ 表示一个基于预训练语言模型的编码器，其参数 $\theta_k \in \mathbb{R}^P$ 在领域 D_k 上进行微调，其中 P 为模型参数总数。给定文本输入 t_u 和 t_i，编码器通过提取最终隐藏状态表示，生成用户表示向量 $\mathbf{r}_u \in \mathbb{R}^d$ 和物品表示向量 $\mathbf{r}_i \in \mathbb{R}^d$：

$$\mathbf{r}_u = f\left(t_u \mid \theta_k\right), \ \mathbf{r}_i = f\left(t_i \mid \theta_k\right), \tag{3}$$

其中 d 表示最终隐藏状态表示的维度。

用户 u 与物品 i 之间的推荐得分 $\hat{y}_{ui}$ 计算为二者表示向量的余弦相似度：

$$\hat{y}_{ui} = \cos(\mathbf{r}_u, \mathbf{r}_i). \tag{4}$$

模型参数 θ_k 使用交叉熵目标进行优化：

$$\theta_k^* = \arg\min_{\theta_k} \sum_{u \in \mathcal{U}_k} \left(\log \hat{y}_{k,ui^+} + \sum_{i^- \in \mathcal{I}_k} \log\left(1 - \hat{y}_{k,ui^-}\right) \right), \tag{5}$$

其中，i^+ 表示用户序列中的下一个物品，i^- 表示领域 $\mathbb{I}_k$ 中除 i^+ 之外的负样本物品。请注意，领域特定参数 θ_k 从预训练基础模型参数 $\theta_{base} \in \mathbb{R}^P$ 初始化。

3 提出的方法：MergeRec

如图2所示，我们设计了***MergeRec*** 框架，以应对不同领域之间无法共享交互数据这一实际限制，这种情形称为数据隔离的跨域序列推荐。MergeRec 由三个关键组件构成：（1）合并初始化，将多个领域专属微调模型整合为一个单一统一模型，以融合不同领域的知识；（2）伪用户数据构建，无需依赖真实用户交互即可合成有意义的合并样本；（3）协同合并优化，联合优化推荐损失和知识蒸馏损失，实现推荐感知的参数整合。通过这一设计，MergeRec 能够有效保留特定领域的协同过滤信号，同时确保在多个领域中的强泛化能力。

表1：跨领域序列推荐问题设置的分类。

设置	无用户/物品需要重叠	隐私保护	数据隔离
传统跨域序列推荐 [1, 3, 16, 17, 27]	✗	✗	✗
隐私保护 CDSR [19, 34, 39, 42, 48]	✗	▲	✗
数据隔离 CDSR（提出的方法）	✓	✓	✓

3.1 数据隔离的跨域序列推荐

我们为跨域序列推荐正式定义了一种全新且现实的场景，称为数据隔离的跨域序列推荐。如表 1 所示，该场景具有两个关键要求：（i）假设各领域之间不存在重叠；（ii）在跨域模型构建过程中禁止访问用户交互数据，这与传统跨域序列推荐形成对比。在数据隔离的跨域序列推荐场景下，用户和物品在不同领域之间都可能完全不相交（例如，对于 $k \neq k'$ 而言，用户为 $\mathbb{U}_k \cap \mathbb{U}_{k'} = \emptyset$，物品为 $\mathbb{I}_k \cap \mathbb{I}_{k'} = \emptyset$），这反映了由独立运营的现实环境。此外，该场景还施加了严格的数据隔离约束：原始交互日志仅可在各自领域内用于训练领域专属模型，不能在领域之间共享。因此，跨域推荐系统的构建必须在不访问任何特定领域交互数据的情况下进行。相反，我们仅假设可以访问 K 领域专属微调模型 $\{\theta_k\}_{k=1}^K$，而用于训练这些模型的数据集则完全无法访问。数据隔离的跨域序列推荐旨在生成一个能够在多个领域中运行的单一通用序列推荐系统。

3.2 合并初始化

问题定义。 $f_{\theta_k}(u_k) \rightarrow \hat{y}_k$ 表示在领域 $D_k = \{\mathbb{U}_k、\mathbb{I}_k\}$ 的私有数据上微调得到的序列推荐模型。对于任意用户交互序列 $u_k \in \mathbb{U}_k$，该模型会输出候选物品的点击概率向量 $\hat{\mathbf{y}}_k \in \mathbb{R}^{|\mathcal{I}_k|}$。不失一般性，我们假设模型参数由 L 层组成，即，$\theta = \left\{\theta^1, \theta^2, ..., \theta^L\right\}$。

任务向量。 任务向量表示将预训练模型适配到特定下游任务 [4, 8, 9, 11, 25, 31, 38, 44, 46, 49, 50] 所需的参数偏移。在我们的情境中，每个任务都对应于特定领域内的推荐。因此，任务向量捕获领域特定知识，使预训练模型能够专门适配该领域。

形式化地说，领域 k 的任务向量 $\boldsymbol{\tau}_k \in \mathbb{R}^P$ 定义为微调模型 θ_k 的参数与

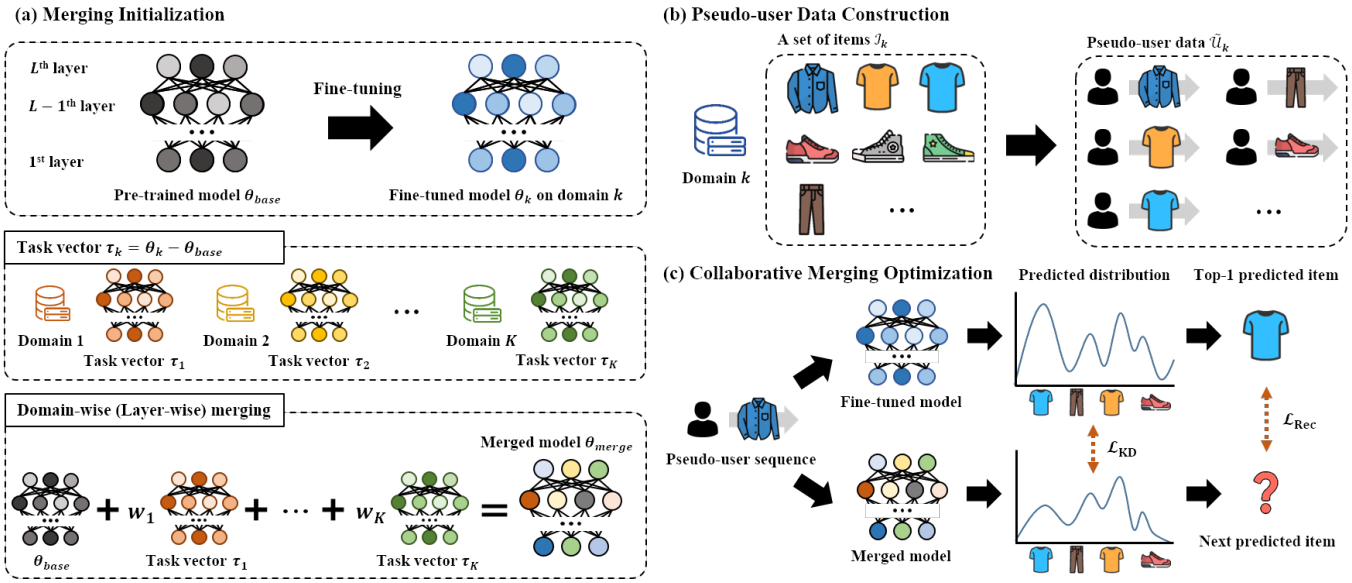

图2：MergeRec 的概览，包含三个主要组成部分。(a) 合并初始化将多领域知识整合到统一模型中。(b) 伪用户数据构建生成单项目序列。(c) 协同合并优化同时优化推荐损失$\mathcal{L}_{Rec}$和蒸馏损失$\mathcal{L}_{KD}$。

原始预训练基础模型 θ_{base}的参数之差：

$$\boldsymbol{\tau}_k = \theta_k - \theta_{base}. \tag{6}$$

其中， P 表示模型参数的总数。

按领域合并。按领域合并通过组合多个微调模型的任务向量，并为每个任务向量赋予特定领域的标量权重 w_k，再将其加到基础模型参数上，从而整合这些模型。合并权重$\mathbf{w} = \{w_1, ..., w_K\}$可以统一设定 [9, 44]，也可以通过自适应学习来反映领域特征 [46]。直观而言，包含更多独特知识的领域可能获得较高权重，而协同过滤信号相似的领域则可能被降低权重。形式上，按领域合并定义如下：

$$\theta_{merge} = \theta_{base} + \sum_{k=1}^{K} w_k \cdot \boldsymbol{\tau}_k. \tag{7}$$

在本研究中，我们以数据驱动的方式学习特定领域权重$\mathbf{w}$。

按层合并。由于深度神经网络的不同层捕获不同抽象层次的特征 [30, 36]，为每个领域应用单一标量权重可能不足以控制跨领域干扰。为实现细粒度集成，我们为每个领域的每一层分配独立的合并权重。

令 $\theta_k = \{\theta_k^1, \ldots, \theta_k^L\}$ 表示领域 k 的微调模型参数。相应的逐层任务向量定义为 $\boldsymbol{\tau}_k = \{\theta_k^1 - \theta_{\text{base}}^1, ..., \theta_k^L - \theta_{\text{base}}^L\}$。逐层合并定义如下：

$$\theta_{merge} = \left\{ \theta_{base}^l + \sum_{k=1}^{K} w_k^l \cdot \tau_k^l \right\}_{l=1}^{L}. \tag{8}$$

层特定合并权重 $\mathbf{w}_k = \{w_k^1, \ldots, w_k^L\}$ 同样以数据驱动的方式进行学习。

3.3 伪用户数据构建

推荐系统中的用户日志通常包含敏感个人信息，无法在领域之间共享，因此难以构建用于学习合并权重的数据。为此，我们提出了一种新颖的伪用户数据构建策略，将领域中的每个物品表示为一个单物品交互序列。我们的设计基于这样一种理念：协同过滤知识封装在领域专属模型中，并且可以在不依赖其训练所用真实用户数据的情况下进行迁移。通过将伪用户用作替代输入，我们的方法能够在无法获取领域专属微调数据的情况下学习合并权重，同时严格保证数据隔离。

形式化地，我们为领域 k 构建伪用户集合如下：

$$\tilde{\mathcal{U}}_k = \{[i] \mid i \in \mathcal{I}_k\}. \tag{9}$$

这些合成样本模拟了各个领域中合理的冷启动用户，因此能够为模型合并提供有意义的信号。尽管每个伪用户序列不包含显式的序列上下文，但它可以作为探针，激发对应领域专属模型 f_{θ_k}中编码的丰富协同过滤知识。因此，我们将每个领域专属模型用作教师模型，其条件分布 $P_{\theta_k}(\cdot \mid [i])$刻画了围绕物品 i的局部共同消费结构，即下一物品似然。通过蒸馏合并模型，使其与这些教师分布保持一致，我们能够有效传递特定领域的协同过滤信号。我们观察到，即使是仅包含单个物品的伪用户序列，也足以实现有效的模型合并（第5节），为协同合并优化提供了实用且隐私保护的基础。虽然将伪用户序列扩展到更长的上下文可能会进一步丰富所传递的信号，但我们将这一有前景的方向留待未来研究。

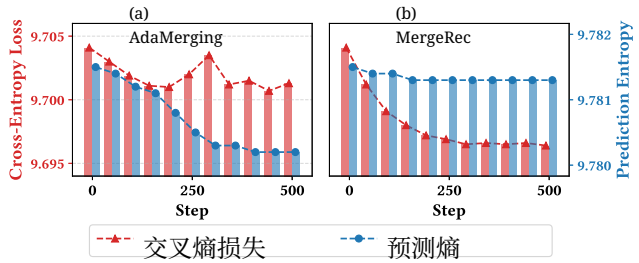

图 3：AdaMerging 和 MergeRec（我们的方法）在训练步数增加过程中的交叉熵损失和预测熵变化。

3.4 协同合并优化

基于熵的优化的不足。 我们从推荐系统的角度考察一种具有代表性的自适应合并方法 AdaMerging [46], 。图3展示了 AdaMerging 和我们提出的 ***MergeRec*** 在八个数据集上，以测试用户序列进行评估时，交叉熵损失和预测熵随训练迭代变化的动态过程。交叉熵损失反映了与推荐目标的一致性，而预测熵则衡量合并模型预测的置信度。如图3(a)所示，AdaMerging 在训练过程中成功降低了预测熵，但未能实现交叉熵损失的相应下降。这一局限源于其对熵最小化的单一关注，这种做法只会放大对排名第一的预测物品的置信度。然而，用户通常表现出多意图 行为模式，而非单一的主导意图 [2, 35, 51], ，因此仅依赖基于熵的优化不足以捕捉领域专属微调模型所学习到的丰富协同过滤信号。

为解决这一问题，我们引入了一种蒸馏损失，通过显式地使合并模型与源自领域专属模型的教师分布对齐，将优化扩展到基于熵的优化之外。MergeRec（图 3（b））同时降低了交叉熵损失和预测熵，表明其能够更一致、更有效地朝着推荐目标进行优化。

联合目标函数。 我们认为，理想的跨域推荐合并模型应同时满足两个基本要求。首先，它应有效捕捉各领域行为模式中体现的多样化用户意图，并检索与这些意图相关的物品。这要求成功地将微调模型中的特定领域协同过滤知识迁移到合并模型中。其次，合并模型应具备较强的判别能力，以准确识别用户在各领域中最有可能点击的物品。为同时满足这些要求，我们提出以下优化函数：

$$\mathcal{L} = \mathcal{L}_{\text{Rec}} + \lambda \cdot \mathcal{L}_{\text{KD}}, \tag{10}$$

其中，λ 是用于平衡两种损失的超参数。在方程（7）和（8）中，我们仅优化合并权重 w，同时固定基础模型参数 θ_{base} 和任务向量 τ。由于仅优化 K 或 $K \times L$ 的特定领域权重，MergeRec 提供了一种计算高效的解决方案。

知识蒸馏损失 L_{KD} 通过使合并模型的预测结果与相应领域专属模型的预测结果保持一致，融合了微调模型中编码的特定领域协同过滤知识。对于领域 k中的伪用户序列 $u \in \tilde{\text{U}}_k$，我们最小化合并模型 θ_{merge} 生成的预测$\hat{y}_{merge}$ 与微调模型 θ_k生成的预测$\hat{y}_k$ 之间的 Kullback-Leibler（KL）散度：

$$\begin{aligned}\mathcal{L}_{\text{KD}} &= \sum_{k=1}^{K} \sum_{u \in \tilde{\mathcal{U}}_k} \text{KL}(\hat{p}_{merge,u} \parallel \hat{p}_{k,u}) \\ &= \sum_{k=1}^{K} \sum_{u \in \tilde{\mathcal{U}}_k} \sum_{i \in \mathcal{I}_k} \hat{p}_{merge,ui} \log \frac{\hat{p}_{merge,ui}}{\hat{p}_{k,ui}},\end{aligned} \tag{11}$$

其中，$\hat{p}_* = \text{softmax}(\hat{y}_*/T)$，$T$表示温度超参数，在我们的实验中根据经验将其设为 1。

推荐损失 L_{Rec} 促使合并模型准确识别符合用户意图的物品：为下一个物品赋予较高分数，并为其他物品赋予较低分数。由于无法获取真实用户序列，我们将伪用户数据输入相应的微调模型，并将由此得到的排名第一的预测物品$\tilde{i}^+$作为正伪标签：

$$\mathcal{L}_{\text{Rec}} = \sum_{k=1}^{K} \sum_{u \in \tilde{\mathcal{U}}_k} \left(\log \hat{y}_{merge,u\tilde{i}^+} + \sum_{i^- \in \mathcal{I}_k} \log (1 - \hat{y}_{merge,ui^-}) \right). \tag{12}$$

4 实验设置

数据集。为模拟跨域推荐环境，我们使用亚马逊数据集[1] [2]中的八个类别：艺术、美容、Instruments、办公用品、食品储藏室、Scientific、体育和玩具。遵循现有工作 [12, 33], ，我们采用5-core设置，即移除交互次数少于五次的用户和物品。详细的数据集统计信息见附录A。

基线方法。 我们将 MergeRec 与以下方法进行比较：

- **零样本**：直接使用预训练文本型序列推荐模型，不针对特定领域进行微调。
- **微调**：使用特定领域的交互数据对预训练模型进行微调。
- **联合学习**：在汇总的多领域数据集上，使用共享参数训练统一模型。
- **任务算术** [9]：通过将任务向量线性相加到预训练模型中，构建跨领域模型。
- **TIES** [44]：通过选择方差较大的参数并对齐其符号，减少任务向量之间的噪声和冲突。
- **AdaMerging** [46]：通过最小化合并模型的预测熵，以无监督方式学习自适应合并权重。

我们在 RecFormer-base/large [13],（一种有代表性的基于文本的序列推荐模型）和 BLaIR-base/large [7],（一种在推荐语料库上经过后训练的语言模型）上评估所有方法，并将其作为骨干网络。为在数据隔离设置下进行公平比较，我们调整 AdaMerging，使其使用与 MergeRec 相同的伪用户数据，并将其视为用于基于熵的优化的无标签输入。实现细节见附录 B。

评估协议。 按照 [12, 33], ，我们采用留一策略划分训练、验证和测试数据集。对于每位

[1]https://cseweb.ucsd.edu/~jmcauley/datasets/amazon/links.html [2] https://cseweb.ucsd.edu/~jmcauley/datasets/amazon_v2/

表2：在四种骨干模型（即 RecFormer-base/large [13] 和 BLaIR-base/large [7]）上与六种基线方法的性能比较。我们报告相对于微调模型性能（设为 100%）的归一化 Recall@10 性能（%）。最佳结果以粗体标出，次佳结果以下划线标出。"*"表示 MergeRec 相较于最佳基线模型取得了统计显著提升（单尾 t 检验的 $p < 0.02$）。

骨干网络	方法	Avg.	Arts	美容	乐器	办公用品	食品储藏室	Sci.	体育	Toys
RecFormer-base	零样本	75.46	76.04	63.66	72.15	61.61	74.23	92.85	71.11	84.94
	联合学习	80.17	77.37	79.53	83.26	73.37	85.34	79.96	69.62	92.94
	权重平均	89.84	91.20	83.05	90.86	77.39	93.62	99.68	**86.66**	93.00
	任务算术	88.95	91.52	81.45	88.12	83.79	92.67	98.90	81.32	82.34
	TIES	91.08	93.29	**85.60**	90.57	**88.22**	92.75	97.68	85.56	86.25
	AdaMerging（按领域）	78.91	87.26	58.66	75.81	64.78	75.02	92.56	67.97	93.11
	AdaMerging（按层）	67.36	67.17	32.92	67.36	58.54	65.64	84.54	65.24	84.93
	MergeRec（按领域）	**92.33***	**96.14***	83.69	90.93	84.27	**95.56***	100.73*	84.00	**93.35***
	MergeRec（按层）	92.08*	95.45*	84.85	**91.13***	83.01	94.86*	**101.44***	85.07	92.14
RecFormer-large	零样本	59.07	72.57	51.89	46.03	39.23	44.34	82.00	62.54	63.57
	联合学习	83.73	83.14	79.48	79.76	73.61	90.27	92.99	91.49	83.61
	权重平均	91.23	92.27	87.10	88.94	79.13	**96.46**	**98.32**	**98.88**	94.36
	任务算术	87.99	91.32	83.32	88.26	82.78	92.08	92.99	87.80	81.14
	TIES	89.96	93.00	85.74	90.25	**88.17**	92.91	94.10	89.09	80.79
	AdaMerging（按领域）	72.59	83.37	59.49	77.55	63.74	61.37	83.96	58.14	70.69
	AdaMerging（按层）	70.80	79.59	52.28	68.27	60.64	67.12	83.08	72.97	72.01
	MergeRec（按领域）	**92.99***	**95.19***	**90.08***	**91.75***	83.39	96.20	97.41	94.74	**96.64***
	MergeRec（按层）	92.50*	94.21*	89.77*	90.25	82.18	96.27	97.86	96.53	96.17*
BLaIR-base	零样本	41.10	45.88	35.57	31.55	27.74	47.20	54.94	34.50	42.32
	联合学习	83.67	82.89	**97.30**	**83.52**	73.09	91.24	74.85	**92.88**	**94.87**
	权重平均	87.90	91.98	93.97	78.41	76.12	**91.39**	**99.95**	89.45	81.81
	任务算术	61.50	57.05	53.43	66.66	61.61	62.42	73.27	49.34	53.78
	TIES	82.95	84.37	78.02	78.26	**87.64**	77.18	88.52	87.26	77.39
	AdaMerging（按领域）	60.60	67.93	52.07	53.40	44.83	50.80	72.94	72.91	73.14
	AdaMerging（按层）	68.48	73.09	61.14	68.86	55.34	57.54	83.92	74.23	70.61
	MergeRec（按领域）	87.40	**94.42***	89.38	77.75	81.44	85.18	95.48	85.05	84.01
	MergeRec（按层）	**88.01**	93.53*	91.83	78.98	81.15	87.82	96.55	87.45	83.06
BLaIR-large	零样本	33.44	38.42	29.99	24.55	19.23	46.48	37.83	27.84	41.70
	联合学习	84.59	83.33	81.01	86.78	78.56	90.60	78.48	83.85	**100.63**
	权重平均	88.99	92.76	89.13	83.82	76.25	93.20	103.56	86.36	81.69
	任务算术	78.86	82.64	79.11	65.86	74.43	83.64	93.56	73.13	67.87
	TIES	90.90	93.78	91.09	87.76	**90.01**	92.58	98.36	**93.67**	75.35
	AdaMerging（按领域）	69.27	82.21	62.36	71.76	51.16	60.90	78.41	77.27	68.45
	AdaMerging（按层）	78.22	84.78	73.80	78.57	62.43	83.01	84.04	71.51	83.00
	MergeRec（按领域）	91.80*	97.80*	93.07*	**91.81***	80.51	89.76	102.88	90.16	82.84
	MergeRec（按层）	**93.70***	**98.73***	**95.40***	91.74*	82.62	**94.17***	**105.18***	90.28	85.63

用户，最近交互的物品用于测试，倒数第二个最近交互的物品用于验证，其余物品用于训练。请注意，训练数据仅用于微调和联合学习。我们使用 Recall@10（R@10）和 NDCG@10（N@10）评估推荐性能。按照 [44],，我们将每种方法的性能除以其对应微调模型的性能进行归一化。归一化后的结果见表2以及图4、图5和图7。

5 实验结果

5.1 总体性能

表2展示了 MergeRec 和七种基线方法在八个数据集和四种骨干模型上的归一化 R@10，其中每个数据集上微调模型的性能均归一化为 100%。相应的归一化 N@10、未归一化 R@10 和 N@10 结果见附录 C。

MergeRec在按领域和逐层变体中，均在所有数据集和骨干模型上始终取得最佳平均性能。具体而言，MergeRec优于Joint

表 3：不同目标领域训练集数据稀疏程度下的性能比较。对于三个数据集（乐器、办公用品和科学）中的每一个，我们将五个源模型（分别在艺术、美容、食品储藏室、体育和玩具数据上训练）与一个目标模型进行合并。每个目标模型均在完整数据集的一个子集（1%、5% 和 10%）上进行训练。“比例”表示用于微调的目标领域训练数据所占的比例，骨干网络采用 RecFormer-base。评估指标为 Recall@10。

比例	方法	Avg.	乐器	办公用品	Sci.
1%	微调	0.0989	0.0745	0.0953	0.1268
	任务算术	0.1069	0.0828	0.0995	0.1385
	TIES	0.0650	0.0158	0.0614	0.1178
	AdaMerging	0.0982	0.0732	0.0898	0.1318
	MergeRec	**0.1089**	**0.0859**	**0.1013**	**0.1394**
5%	微调	0.1057	0.0752	0.1094	0.1325
	任务算术	0.1101	0.0859	0.1050	0.1392
	TIES	0.0800	0.0812	0.1017	0.0570
	AdaMerging	0.1047	0.0814	0.1023	0.1304
	MergeRec	**0.1128**	**0.0888**	**0.1098**	**0.1399**
10%	微调	0.1118	0.0838	**0.1175**	0.1341
	任务算术	0.1107	0.0860	0.1055	**0.1405**
	TIES	0.0840	0.0809	0.1017	0.0693
	AdaMerging	0.0982	0.0755	0.0935	0.1258
	MergeRec	**0.1120**	**0.0887**	0.1088	0.1386

与 AdaMerging 相比，增益分别为 8.72% 和 17.21%。这表明 MergeRec 能够同时增强合并模型的排序判别能力，并有效迁移微调模型中的领域特定协同过滤知识。与此同时，AdaMerging 在所有数据集和骨干模型上的表现都明显较差，这说明仅仅放大预测置信度不足以捕获多个领域中存在的多样化协同过滤信号。此外，MergeRec 的表现优于无需训练的模型合并方法（即Task Arithmetic 和 TIES），平均增益分别为 9.90% 和 3.04%。

若干模型合并方法（即MergeRec、权重平均和 TIES）的表现始终优于 JointLearning。这些结果表明，模型合并能够有效捕获互补的领域知识，并在不依赖跨领域训练数据的情况下提升推荐质量。这进一步凸显了模型合并的实际优势：它不仅能降低计算开销，还能在无需端到端重新训练的情况下实现跨领域的协同知识迁移。

跨领域合并对数据稀缺领域尤其有益。在 Scientific 数据集上，与各自微调后的模型相比，MergeRec 使 RecFormer-base（按领域）的性能提升了 1.44%，使 BLaIR-large（逐层）的性能提升了 5.18%。这些提升归因于 Scientific 领域的用户数量有限，因此合并模型能更充分地受益于从其他领域迁移来的跨域知识。

总体而言，这些结果表明，MergeRec 为推荐系统中的跨领域模型合并提供了一种稳健且可扩展的解决方案，能够在不同领域和骨干架构中持续带来显著的性能提升。

表 4：使用 RecFormer-base 骨干网络的基于任务向量的模型合并方法在未见领域上的性能比较。我们在艺术、美容、食品储藏室、体育和玩具数据集上训练合并模型，并在乐器、办公用品和科学数据集上进行测试。评估指标为 Recall@10。

方法	Avg.	乐器	办公用品	Sci.
任务算术	0.1062	0.0817	0.0984	0.1385
TIES	0.1062	0.0817	0.0988	0.1380
AdaMerging	0.0998	0.0759	0.0901	0.1333
MergeRec	**0.1081**	**0.0849**	**0.1002**	**0.1393**

5.2 数据稀缺情况下的模型合并

在推荐系统发展的早期阶段，收集充足的数据往往颇具挑战。在这种数据稀缺条件下，模型合并可以利用数据丰富领域中的知识，为提升模型的泛化能力提供一种有前景的解决方案。我们研究了模型合并能否提升数据有限领域中的推荐性能。

为模拟这一场景，我们将八个领域分为两组：五个源领域（艺术、美容、食品储藏室、体育、玩具）和三个目标领域（Instruments、办公用品、Scientific）。随后，我们通过从全部用户中随机抽取k%的用户（$k = 1$，5，10）来构建数据稀缺的训练集，以此改变目标领域的数据稀缺程度。接着，将每个在数据稀缺的目标域上训练的微调模型，与在五个源领域上训练的微调模型合并。

表 3 使用 RecFormer-base 骨干网络，在不同程度的数据稀缺情况下，将 MergeRec 与五种合并方法进行了比较。MergeRec 始终优于相应的微调模型，表明即使在极端数据稀缺的情况下，它也具备强大的跨领域知识迁移能力。此外，在所有数据稀缺程度下，MergeRec 均持续优于 AdaMerging，这说明我们面向推荐的优化能够比熵最小化方法更有效地捕获可迁移的协同过滤模式。这些结果证实，MergeRec 能够可靠地将协同过滤信号从数据丰富的领域迁移到数据稀缺的领域。

5.3 在未见领域上的性能

为考察在目标域没有任何交互数据这一极端条件下模型合并是否仍然有效，我们合并了在五个源领域（艺术、美容、食品储藏室、体育、玩具）上训练的模型，并在三个未见目标领域（乐器、办公用品、科学）上评估其性能。需要注意的是，在构建合并模型时，既未使用来自目标领域的交互数据，也未使用目标领域的领域专属模型。

表 4 展示了四种合并方法在未见领域上的性能。MergeRec 始终优于其他方法，相较于任务算术的提升最高可达 3.92%。这些结果证明了 MergeRec 强大的泛化能力，并凸显了其在不依赖任何目标领域数据的情况下，将知识迁移到完全未见领域的有效性。

5.4 进一步分析

合并领域数量的影响。为评估 MergeRec 的稳健性，我们分析了推荐性能如何随合并模型数量变化，如图 4 所示。

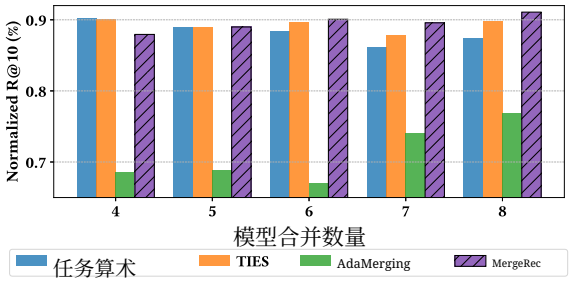

图 4：模型合并在不同数据集数量下的归一化平均性能。

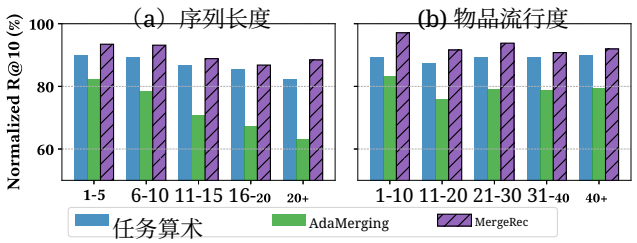

图 5：八个数据集上用户组和物品组的归一化平均性能。用户组和物品组分别按序列长度和物品流行度划分。

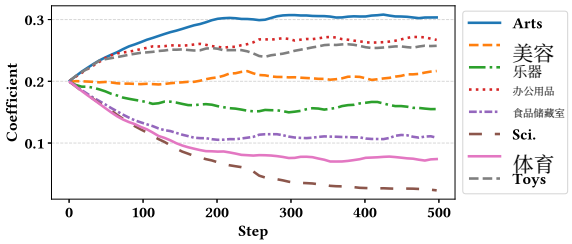

图6：RecFormer-base 骨干网络上领域特定系数随训练步数的变化。

任务算术和 TIES 在合并少量领域（四个或五个）时表现出较强的竞争力，但随着纳入的领域增多，其性能会逐渐下降，最终落后于 MergeRec。AdaMerging 在所有设置下的表现始终不佳，这表明其基于熵的优化策略无法有效适应推荐任务。相比之下，无论合并领域的数量如何，MergeRec 都能保持稳定的性能。值得注意的是，随着更多领域的整合，其平均性能有所提升，凸显了它有效整合并利用日益增多的多样化领域知识的强大能力。

不同用户组和物品组上的性能。为进一步分析性能提升的来源，我们根据用户历史记录长度和目标物品流行度将测试数据划分为多个组。图 5 展示了三种合并方法在五个序列长度组和五个物品流行度范围内的性能。所有方法均在八个数据集上进行评估，结果取平均值，并按各组中相应微调模型的性能进行归一化。

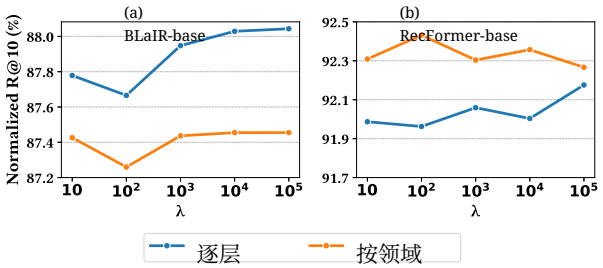

图7：不同 λ下的归一化平均性能。

如图 5(a) 所示，随着序列长度增加，AdaMerging 的性能严重下降。这一结果表明，AdaMerging 难以捕获较长序列中的协同过滤信号，因为它在优化过程中没有学习物品间关系。相比之下，MergeRec 通过所提出的目标函数有效迁移物品间关系，在所有序列长度下始终取得最佳性能。图 5(b) 显示，MergeRec 在所有物品流行度范围内均优于其他方法。值得注意的是，在流行度最低的物品组（1–10）中，MergeRec 相较于其他方法取得了显著的性能提升，凸显了其利用跨域知识推荐较不热门物品的能力。

领域特定权重的变化。为了解模型在合并过程中如何适应不同领域，我们使用 RecFormer-base 模型分析了八个领域中逐域合并权重 $\mathbf{w}$ 的训练轨迹。图6展示了各领域权重在训练过程中的变化。我们观察到，权重逐渐收敛到各领域不同的取值，这表明模型学会了区分各个领域的相对重要性。值得注意的是，规模较大的领域（例如艺术和办公用品）会收敛到较高的权重系数。这些结果表明，具有更复杂协同过滤模式的领域会得到更高的重视，使合并模型能够更好地保留领域特定知识。其他骨干模型的相应结果见附录D。

5.5 超参数敏感性

图7展示了权衡超参数 λ 对推荐性能的影响。结果是对全部八个数据集取平均，并根据相应微调模型的性能进行了归一化。我们发现，随着 λ 增加，性能通常会有所提升，但使用 RecFormer-base 骨干网络进行逐域合并时除外。这表明，通常更加重视知识蒸馏损失会更有效。这可能是因为模型从特定领域教师模型中学习完整的物品预测分布，使合并模型能够捕捉更丰富、更有信息量的协同过滤信号。相比之下，仅专注于提升 Top-1 预测置信度的方法所提供的监督信号则较为有限。

6 相关工作

6.1 跨域序列推荐跨域序列推荐 [3] 已成为缓解数据稀疏性和单领域推荐系统中冷启动问题的有效方法

系统 [6, 12, 14, 15, 33]。现有的跨域序列推荐方法可分为三类：单目标跨域序列推荐、双目标跨域序列推荐和多目标跨域序列推荐。此外，在注重隐私保护和联邦学习的场景中，隐私保护跨域推荐（CDR）日益受到关注。

单目标跨域序列推荐。单目标跨域序列推荐 [1, 16, 17, 20] 是研究最为广泛的设置，旨在通过从拥有大量用户-物品交互或辅助信息的数据丰富的源域中迁移知识，提升数据稀缺的目标域中的推荐性能。C^{2}DSR [1] 通过图神经网络和自注意力机制联合建模序列内和序列间的物品关系。为增强自注意力模块，MAN [16] 引入局部和全局注意力模块，以捕捉领域特定信息和跨领域信息。近期，LLM4CDSR [17] 利用大语言模型从文本属性中生成物品语义表示，并从交互序列中生成分层用户画像，从而促进跨域知识迁移。

双目标与多目标跨域序列推荐。双目标与多目标跨域序列推荐 [26,27, 37, 52] 旨在同时提升多个领域中的推荐性能。DTCDR [52] 采用多任务学习框架，通过共享嵌入在多个领域之间双向迁移用户偏好表示。近期研究 [26, 27] 凸显了负迁移问题，即来自其他领域的知识可能无法带来有益贡献。为解决这一问题，CGRec [26] 和 SyNCRec [27] 根据单领域与跨域序列推荐任务之间估计的迁移差距，引入自适应损失加权。然而，这些方法从根本上依赖重叠用户或物品来实现知识迁移，并未考虑现实场景中通常要求的隐私约束。

隐私保护跨域推荐推荐 [19, 34, 39, 42,48] 旨在改善无法访问或完全无法获取各个领域原始数据时的推荐性能。早期研究 [42] 将个性化组件与可迁移组件分离，以实现符合隐私规范的推荐。P2M2-CDR [39] 将领域共有嵌入与领域特定嵌入解耦，同时采用本地差分隐私对共享表示进行扰动。FedGCDR [48] 进一步采用联邦图学习，并结合基于差分隐私的知识提取和图扩展，以缓解负迁移问题。然而，这些方法通常依赖跨领域的重叠用户，并要求训练期间进行跨域协调；在此过程中，领域特定数据会与其他领域的数据一同使用（例如，交换模型参数或梯度）。因此，它们并未完全满足数据隔离 约束。

在本文中，我们率先将模型合并应用于数据隔离的多目标跨域序列推荐，为集成式多领域推荐提供了一种可扩展的解决方案，无需直接访问领域特定用户交互数据，从而保护用户隐私。

6.2 模型合并

模型合并 [4, 5, 8–11, 18, 21, 25, 31, 32, 38, 40, 43–46, 49, 50] 旨在通过整合多个在不同领域或任务上训练的模型所掌握的知识，提升领域特定微调模型的泛化能力，通常无需访问训练数据。最简单的方法 [10, 40] 是对共享同一骨干预训练模型的微调模型进行参数平均。

除了这些早期方法外，基于任务向量的模型合并 [4, 8, 9, 11, 25, 31, 38, 44, 46, 49, 50] 也已被提出，以实现更有效的知识整合。任务向量定义为各微调模型与预训练模型之间的参数差异。它们可以被理解为编码领域或任务特定适配的方向。任务算术 [9]表明，可以组合任务向量来构建多任务模型，也可以将其取反以减弱（或消除）任务特定知识。TIES [44]会选择因任务而发生较大变化的参数，并解决符号冲突，从而减少干扰。AdaMerging [46]通过在无标签测试样本上进行熵最小化，学习领域或逐层的合并权重，从而扩展了任务向量的线性组合。尽管模型合并已在计算机视觉和自然语言处理中展现出显著成效，但其在推荐系统中的潜力仍基本未得到探索。

7 结论

在本研究中，我们针对现实约束下现有跨域序列推荐的根本局限展开研究：用户交互数据无法在不同领域间共享。为此，我们提出了***MergeRec***，这是一个新颖的框架，将基于任务向量的模型合并应用于一种称为数据隔离的跨域序列推荐的新问题设定。MergeRec 包含三个关键组件：(1) 合并初始化 通过基于免训练任务向量的合并来构建初始合并模型。(2) 伪用户数据构建 从领域物品合成虚拟交互序列，从而无需暴露敏感用户数据即可提取协同过滤信号。(3) 协同合并优化 联合优化推荐损失和知识蒸馏损失，在保留排序效果的同时，促进领域特定协同过滤模式的迁移。大量实验表明，MergeRec 始终优于现有模型合并基线方法，并显著提升了泛化能力，包括在未见领域和数据稀缺领域上的表现，凸显了模型合并作为一种可扩展的隐私保护解决方案，在构建通用推荐系统方面的潜力。

数据的伦理使用

本文使用公开可用的数据集（亚马逊产品评论数据集），其中不包含任何个人身份信息，因此无需获得机构伦理审查委员会的批准。所有实验均使用为学术研究目的发布的匿名化交互序列。

致谢

本研究由韩国政府（MSIT）资助的信息与通信技术规划评估院（IITP）项目和韩国国家研究基金会（NRF）项目支持（编号：IITP-RS-2019-II190421、IITP-RS-2022-II220680、RS-2025-25442569、NRF-RS-2025-00564083；各项目对本研究的资助比例均为 25%）。

参考文献

[1] Jiangxia Cao、Xin Cong、Jiawei Sheng、Tingwen Liu 和 Bin Wang。2022。对比式跨域序列推荐。发表于 CIKM。138–147。[2] Yukuo Cen、Jianwei Zhang、Xu Zou、Chang Zhou、Hongxia Yang 和 Jie Tang。2020。可控多兴趣推荐框架。发表于 KDD。2942–2951。[3] Shu Chen、Zitao Xu、Weike Pan、Qiang Yang 和 Zhong Ming。2024。跨域序列推荐综述。发表于 IJCAI。7989–7998。[4] Guodong Du、Junlin Lee、Jing Li、Runhua Jiang、Yifei Guo、Shuyang Yu、Hanting Liu、Sim Kuan Goh、Ho-Kin Tang、Daojing He 和 Min Zhang。2024。模型合并的参数竞争平衡。发表于 NeurIPS。[5] Antonio Andrea Gargiulo、Donato Crisostomi、Maria Sofia Bucarelli、Simone Scardapane、Fabrizio Silvestri 和 Emanuele Rodolà。2025。任务奇异向量：减少模型合并中的任务干扰。发表于 CVPR。18695–18705。[6] Balázs Hidasi、Alexandros Karatzoglou、Linas Baltrunas 和 Domonkos Tikk。2016。基于循环神经网络的会话推荐。发表于 ICLR。[7] Yupeng Hou、Jiacheng Li、Zhankui He、An Yan、Xiusi Chen 和 Julian McAuley。2024。连接语言与物品，实现检索与推荐。arXiv 预印本 arXiv:2403.03952（2024）。[8] Chenyu Huang、Peng Ye、Tao Chen、Tong He、Xiangyu Yue 和 Wanli Ouyang。2024。EMR-Merging：无需调优的高性能模型合并。发表于 NeurIPS。[9] Gabriel Ilharco、Marco Túlio Ribeiro、Mitchell Wortsman、Ludwig Schmidt、Hannah Hajishirzi 和 Ali Farhadi。2023。使用任务算术编辑模型。发表于 ICLR。[10] Gabriel Ilharco、Mitchell Wortsman、Samir Yitzhak Gadre、Shuran Song、Hannah Hajishirzi、Simon Kornblith、Ali Farhadi 和 Ludwig Schmidt。2022。通过权重插值修补开放词汇模型。发表于 NeurIPS。[11] Ruochen Jin、Bojian Hou、Jiancong Xiao、Weijie J. Su 和 Li Shen。2025。仅微调注意力模块：增强任务算术中的权重解耦。发表于 ICLR。[12] Wang-Cheng Kang 和 Julian J. McAuley。2018。自注意力序列推荐。发表于 ICDM。197–206。[13] Jiacheng Li、Ming Wang、Jin Li、Jinmiao Fu、Xin Shen、Jingbo Shang 和 Julian J. McAuley。2023。文本即所需的一切：学习用于序列推荐的语言表示。发表于 KDD。1258–1267。[14] Muyang Li、Zijian Zhang、Xiangyu Zhao、Wanyu Wang、Minghao Zhao、Runze Wu 和 Ruocheng Guo。2023。AutoMLP：用于序列推荐的自动化多层感知机。发表于 WWW。ACM，1190–1198。[15] Muyang Li、Xiangyu Zhao、Chuan Lyu、Minghao Zhao、Runze Wu 和 Ruocheng Guo。2022。MLP4Rec：用于序列推荐的纯多层感知机架构。发表于 IJCAI。ijcai.org，2138–2144。[16] Guanyu Lin、Chen Gao、Yu Zheng、Jianxin Chang、Yanan Niu、Yang Song、Kun Gai、Zhiheng Li、Depeng Jin、Yong Li 和 Meng Wang。2024。用于跨域序列推荐的混合注意力网络。发表于 WSDM。405–413。[17] Qidong Liu、Xiangyu Zhao、Yejing Wang、Zijian Zhang、Howard Zhong、Chong Chen、Xiang Li、Wei Huang 和 Feng Tian。2025。连接领域：大语言模型增强的跨域序列推荐。发表于 SIGIR。1582–1592。[18] Zhenyi Lu、Chenghao Fan、Wei Wei、Xiaoye Qu、Dangyang Chen 和 Yu Cheng。2024。Twin-Merging：模型合并中模块化专业知识的动态整合。发表于 NeurIPS 2024。[19] Ziang Lu、Lei Guo、Xu Yu、Zhiyong Cheng、Xiaohui Han 和 Lei Zhu。2025。用于隐私保护跨域推荐的联邦语义学习。CoRR abs/2503.23026（2025）。[20] Muyang Ma、Pengjie Ren、Zhumin Chen、Zhaochun Ren、Lifan Zhao、Peiyu Liu、Jun Ma 和 Maarten de Rijke。2022。用于跨域序列推荐的混合信息流。ACM 知识发现与数据挖掘汇刊 16，4（2022），64:1–64:32。[21] Michael Matena 和 Colin Raffel。2022。使用费舍尔加权平均合并模型。发表于 NeurIPS。[22] Jaewan Moon、Yoonki Jeong、Dong-Kyu Chae、Jaeho Choi、Hyunjung Shim 和 Jongwuk Lee。2023。CoMix：面向隐式数据集的混合增强协同过滤。信息科学 628（2023），254–268。[23] Jaewan Moon、Hye-young Kim 和 Jongwuk Lee。2023。已经足够：放宽推荐线性自编码器中的对角约束。发表于 SIGIR。1639–1648。[24] Jaewan Moon、Seongmin Park 和 Jongwuk Lee。2025。大语言模型增强的线性自编码器用于推荐。发表于 CIKM。5036–5040。[25] Guillermo Ortiz-Jiménez、Alessandro Favero 和 Pascal Frossard。2023。切空间中的任务算术：改进预训练模型的编辑。发表于 NeurIPS 2023。[26] Chung Park、Taesan Kim、Taekyoon Choi、Junui Hong、Yelim Yu、Mincheol Cho、Kyunam Lee、Sungil Ryu、Hyungjun Yoon、Minsung Choi 和 Jaegul Choo。2023。破解负迁移之谜：一种用于跨域序列推荐的合作博弈论方法。发表于 CIKM。2024–2033。

[27] Chung Park、Taesan Kim、Hyungjun Yoon、Junui Hong、Yelim Yu、Mincheol Cho、Minsung Choi 和 Jaegul Choo。2024。Pacer and Runner：单域与跨域序列推荐之间的协作学习框架。载于 SIGIR。2071–2080。[28] Seongmin Park、Mincheol Yoon、Jae-woong Lee、Hogun Park 和 Jongwuk Lee。2023。深入理解协同过滤损失函数。载于 CIKM。2034–2043。[29] Seongmin Park、Mincheol Yoon、Hye young Kim 和 Jongwuk Lee。2025。线性推荐器为何需要归一化？载于 SIGIR。2142–2151。[30] Anna Rogers、Olga Kovaleva 和 Anna Rumshisky。2020。BERTология入门：关于 BERT 工作原理的已知内容。计算语言学协会汇刊 8（2020），842–866。[31] Daiki Shirafuji、Makoto Takenaka 和 Shinya Taguchi。2025。偏置向量：用任务算术方法缓解语言模型中的偏差。载于 COLING。2799–2813。[32] George Stoica、Daniel Bolya、Jakob Bjorner、Pratik Ramesh、Taylor Hearn 和 Judy Hoffman。2024。ZipIt!：无需训练即可合并不同任务的模型。载于 ICLR。[33] Fei Sun、Jun Liu、Jian Wu、Changhua Pei、Xiao Lin、Wenwu Ou 和 Peng Jiang。2019。BERT4Rec：基于变换器双向编码器表示的序列推荐。载于 CIKM。1441–1450。[34] Changxin Tian、Yuexiang Xie、Xu Chen、Yaliang Li 和 Xin Zhao。2024。基于联邦图学习的隐私保护跨域推荐。《ACM 信息系统汇刊》 42, 5 2024 135 1 135 29 10 1145 3653448（），：–：。[35] 2022 https://doi.org/。/ Yu Tian、Jianxin Chang、Yanan Niu、Yang Song 和 Chenliang Li。。当多层级遇上多兴趣：用于序列推荐的多粒度神经模型。载于 SIGIR。1632–1641。[36] Betty van Aken、Benjamin Winter、Alexander Löser 和 Felix A. Gers。2019。BERT 如何回答问题？变换器表示的逐层分析。载于 CIKM。1823–1832。[37] Hao Wang、Mingjia Yin、Luankang Zhang、Sirui Zhao 和 Enhong Chen。2025。MF-GSLAE：用于双目标跨域推荐的多因素用户表示预训练框架。《ACM 信息系统汇刊》43，2（2025），30:1–30:28。[38] Ke Wang、Nikolaos Dimitriadis、Guillermo Ortiz-Jiménez、François Fleuret 和 Pascal Frossard。2024。定位任务信息以改进模型合并与压缩。载于 ICML。[39] Li Wang、Lei Sang、Quangui Zhang、Qiang Wu 和 Min Xu。2024。用于跨域推荐的多模态数据隐私保护框架。Knowledge-Based Systems 304（2024），112529。[40] Mitchell Wortsman、Gabriel Ilharco、Jong Wook Kim、Mike Li、Simon Kornblith、Rebecca Roelofs、Raphael Gontijo Lopes、Hannaneh Hajishirzi、Ali Farhadi、Hongseok Namkoong 和 Ludwig Schmidt。2022。零样本模型的稳健微调。载于 CVPR。7949–7961。[41] Chuhan Wu、Fangzhao Wu、Yang Cao、Yongfeng Huang 和 Xing Xie。2021。FedGNN：用于隐私保护推荐的联邦图神经网络。CoRR abs/2102.04925（2021）。[42] Meihan Wu、Li Li、Chang Tao、Eric Rigall、Xiaodong Wang 和 Cheng-Zhong Xu。2022。FedCDR：用于隐私保护评分预测的联邦跨域推荐。载于 CIKM。2179–2188。[43] Zhengqi Xu、Ke Yuan、Huiqiong Wang、Yong Wang、Mingli Song 和 Jie Song。2024。免训练预训练模型合并。载于 CVPR。5915–5925。[44] Prateek Yadav、Derek Tam、Leshem Choshen、Colin A. Raffel 和 Mohit Bansal。2023。TIES-Merging：解决模型合并时的干扰。载于 NeurIPS。[45] Enneng Yang、Li Shen、Zhenyi Wang、Guibing Guo、Xiaojun Chen、Xingwei Wang 和 Dacheng Tao。2024。面向多任务模型合并的表示手术。载于 ICML。[46] Enneng Yang、Zhenyi Wang、Li Shen、Shiwei Liu、Guibing Guo、Xingwei Wang 和 Dacheng Tao。2024。AdaMerging：用于多任务学习的自适应模型合并。载于 ICLR。[47] Liu Yang、Ben Tan、Vincent W. Zheng、Kai Chen 和 Qiang Yang。2020。联邦推荐系统。载于《联邦学习——隐私与激励》。第12500卷。Springer，225–239。[48] Ziqi Yang、Zhaopeng Peng、Zihui Wang、Jianzhong Qi、Chaochao Chen、Weike Pan、Chenglu Wen、Cheng Wang 和 Xiaoliang Fan。2024。用于跨域推荐的联邦图学习。载于 NeurIPS。[49] Kotaro Yoshida、Yuji Naraki、Takafumi Horie、Ryosuke Yamaki、Ryotaro Shimizu、Yuki Saito、Julian J. McAuley 和 Hiroki Naganuma。2025。掌握任务算术： τJp 是权重解耦的关键指标。载于 ICLR。[50] Frederic Z. Zhang、Paul Albert、Cristian Rodriguez Opazo、Anton van den Hengel 和 Ehsan Abbasnejad。2024。利用任务向量和学习到的各向异性缩放进行知识组合。载于 NeurIPS 2024。[51] Shengyu Zhang、Lingxiao Yang、Dong Yao、Yujie Lu、Fuli Feng、Zhou Zhao、Tat-Seng Chua 和 Fei Wu。2022。Re4：学习针对多兴趣推荐进行再对比、再关注和再构建。载于 WWW。2216–2226。[52] Feng Zhu、Chaochao Chen、Yan Wang、Guanfeng Liu 和 Xiaolin Zheng。2019。DTCDR：双目标跨域推荐框架。载于 CIKM。1533–1542。

表5：数据集统计，包括用户、物品、交互数量及密度。

数据集	# 用户	# 物品	# Inter.	密度
Arts	56,210	22,855	492,492	0.04%
美容	22,363	12,101	198,502	0.07%
乐器	27,530	10,611	231,312	0.08%
办公用品	101,499	27,932	798,912	0.03%
食品储藏室	14,180	4,968	137,769	0.20%
Sci.	11,041	5,327	76,896	0.13%
体育	35,598	18,357	296,337	0.05%
Toys	19,412	11,924	167,597	0.07%

A 数据集统计

表5总结了各领域的统计信息，包括用户数、物品数、交互数和数据集密度。密度计算方式如下：$\#$ 交互数 $\#$ 用户数$\times\#$ 物品数。

B 实现细节

所有方法，包括 MergeRec 和基线方法，均使用 PyTorch 实现。对于 RecFormer-base [13],，我们使用官方预训练检查点[3],；对于 RecFormer-large，我们则按照原论文中描述的协议对模型进行预训练。对于 BLaIR-base[4]和 BLaIR-large[5] [7],，我们使用 HuggingFace 上提供的官方预训练检查点。微调采用批内负采样，批次大小为 64。对于合并基线方法，我们采用其在无验证集设置下的超参数配置，即不使用训练、验证或测试数据。具体而言，对于任务算术，我们设置 $w_1 = w_2 = \cdots = w_K = 0.4$ [9],；对于 TIES，我们设置 $w = 1$ [44]。TIES 使用排名前 20% 的参数。对于 AdaMerging [46] 和 MergeRec，所有系数均初始化为 0.2，并使用 Adam优化器以 0.001 的学习率和 16 的批次大小优化 500 台阶。对于 MergeRec，我们设置 $\lambda = 1{,}000$，以平衡两个损失函数的尺度。所有报告的性能指标均为五个随机种子下计算所得的平均值。在显著性检验中，我们假设确定性合并方法（即权重平均、任务算术和 TIES）在五次运行中的性能值相同。

C 其他指标的整体表现

表6展示了 MergeRec 和七种基线方法在四种骨干网络模型上的八个数据集上的归一化 N@10，其中在每个数据集上训练的微调模型的性能设为 100%。此外，我们分别在表7和表8中报告了未归一化 R@10 和 N@10 的结果。我们观察到，N@10（表6）的趋势与 R@10（表2）相似。

(i) MergeRec 在所有数据集和骨干模型上的平均表现均达到最佳。这表明 MergeRec 能够有效迁移领域特定知识并提升模型判别能力，而 AdaMerging 在各方面的效果都较为逊色。**(ii)** MergeRec、权重平均和 TIES 等合并方法始终

[3]https://github.com/AaronHeee/RecFormer [4] https://huggingface.co/hyp1231/blair-roberta-base [5] https://huggingface.co/hyp1231/blair-roberta-large

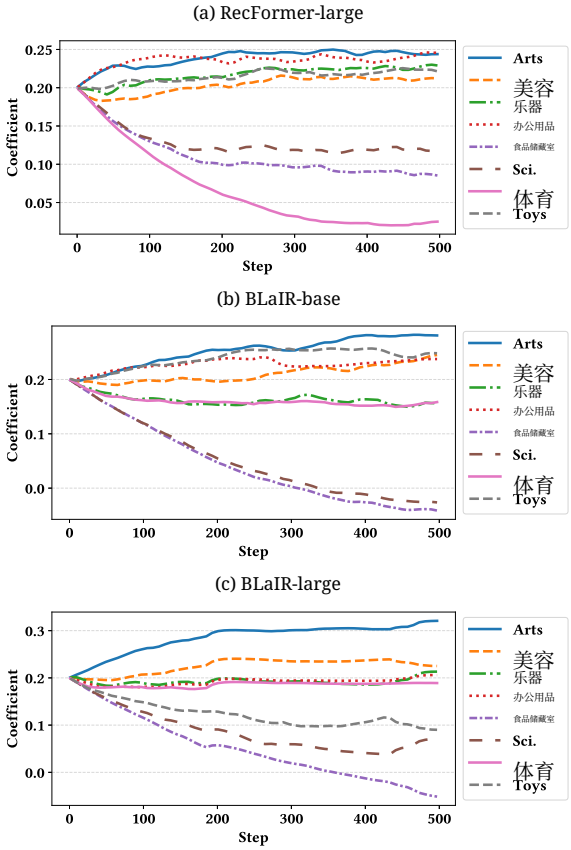

图8：RecFormer-large 和 BLaIR-base/large 骨干网络在训练步数增加过程中的领域特定系数动态。

优于联合学习，表明参数合并能够在没有跨领域训练数据的情况下，高效地跨领域迁移知识并降低计算成本。**(iii)** 跨领域合并对于数据有限的领域尤其有帮助，例如 Scientific。在这些领域中，得益于更有效的知识迁移，MergeRec 相较于微调模型展现出显著提升。

D 其他骨干网络上的域特定权重动态

我们进一步分析了剩余三个骨干模型——RecFormer-large 和 BLaIR-base/large——中特定领域权重w 的演变情况。与我们对 RecFormer-base 的观察一致，我们发现，数据规模较大的领域（例如，艺术 和 办公用品）往往会收敛到较高的权重系数。这一趋势表明，这些模型也倾向于捕捉大规模领域中更复杂的协同模式，从而分配更多表征能力来保留领域特定知识。与这一普遍趋势不同，BLaIR 模型中 Pantry 数据集的合并权重收敛为负值。这种现象可能归因于 Pantry 数据集中缺少类别信息；该数据集与其他领域有所不同，这可能导致模型合并过程中出现不对齐。

表6：在四种骨干模型上与六种基线方法的性能比较，即 RecFormer-base/large [13] 和 BLaIR-base/large [7]。我们报告归一化 NDCG@10 性能（%），其中微调模型的性能为 100%。最佳结果以粗体标出，次佳结果以下划线标出。“*”表示 MergeRec 相较于最佳基线模型取得了统计显著提升（单尾 t 检验的 $p < 0.02$）。

骨干网络	方法	Avg.	Arts	美容	乐器	办公用品	食品储藏室	Sci.	体育	Toys
RecFormer-base	零样本	67.52	63.59	61.57	55.88	52.63	69.33	92.79	70.75	81.44
	联合学习	73.05	67.41	75.48	76.13	66.69	82.75	72.12	69.55	**87.62**
	权重平均	83.15	82.88	76.42	75.67	70.29	92.37	99.68	**82.70**	86.20
	任务算术	84.08	86.47	75.99	76.05	76.37	91.85	99.52	78.08	75.53
	TIES	85.60	87.16	**79.61**	76.52	**81.25**	91.95	97.59	80.72	80.11
	AdaMerging（按领域）	71.17	75.47	55.19	60.35	56.45	71.09	90.67	67.83	87.15
	AdaMerging（按层）	61.61	60.94	32.01	52.35	51.36	64.14	82.60	64.32	79.22
	MergeRec（按领域）	**86.07***	**87.31**	76.54	**79.14***	76.87	**94.50***	100.48*	80.36	84.97
	MergeRec（按层）	85.37	86.18	77.32	77.77*	75.39	93.82*	**101.00***	81.55	83.87
RecFormer-large	零样本	53.55	65.05	50.33	34.05	31.26	40.96	82.63	61.27	63.58
	联合学习	75.76	72.09	75.48	66.50	66.67	86.94	87.67	85.34	78.60
	权重平均	85.84	86.37	81.64	76.44	74.60	94.63	98.70	**91.35**	88.87
	任务算术	85.70	93.60	80.00	77.58	77.38	91.60	94.56	81.75	77.75
	TIES	87.64	94.09	81.66	80.60	**82.07**	92.36	96.49	83.51	76.81
	AdaMerging（按领域）	67.29	77.00	56.84	61.87	56.92	55.66	82.95	56.31	69.72
	AdaMerging（按层）	64.96	71.01	51.70	53.68	54.66	62.19	81.38	69.86	69.99
	MergeRec（按领域）	**89.40***	**96.27***	84.65*	**80.68**	78.21	94.07	98.97	88.13	90.48*
	MergeRec（按层）	88.62*	93.40	**85.08***	78.27	77.58	**94.80**	**99.58**	89.95	**90.66***
BLaIR-base	零样本	33.06	32.86	33.22	23.98	20.80	40.87	45.76	34.05	40.78
	联合学习	78.48	79.94	**97.16**	**79.19**	67.59	**88.03**	68.00	**92.07**	**92.08**
	权重平均	78.13	78.59	85.53	67.01	64.70	85.65	**94.62**	82.03	75.85
	任务算术	55.37	50.11	51.82	55.54	52.30	58.80	66.56	49.09	51.86
	TIES	74.72	72.17	73.04	65.61	**78.62**	71.78	82.47	81.81	73.44
	AdaMerging（按领域）	50.28	51.53	49.02	41.46	35.38	43.17	66.06	68.61	69.18
	AdaMerging（按层）	57.91	56.90	56.93	53.60	44.83	49.51	77.66	69.20	65.29
	MergeRec（按领域）	78.30	**81.33***	81.78	65.59	70.51	78.22	92.14	79.46	78.36
	MergeRec（按层）	**79.18**	81.23*	84.30	66.94	70.26	81.88	93.35	81.73	77.47
BLaIR-large	零样本	26.85	27.51	27.83	18.08	14.22	41.96	32.15	28.46	38.78
	联合学习	81.67	81.00	80.58	**85.57**	75.84	**92.12**	71.85	85.25	**97.07**
	权重平均	81.50	83.19	83.57	71.84	67.74	90.68	98.42	84.70	76.31
	任务算术	73.00	73.33	77.37	57.23	66.13	81.75	89.62	75.40	65.31
	TIES	85.74	85.71	87.92	76.38	**83.56**	89.96	97.93	**90.85**	72.04
	AdaMerging（按领域）	60.21	68.38	58.18	57.61	42.43	56.16	72.01	75.84	62.50
	AdaMerging（按层）	70.67	74.10	69.21	66.94	54.57	80.67	79.05	70.79	77.29
	MergeRec（按领域）	84.46	86.73*	85.69	78.68	72.36	87.45	100.83*	86.61	77.22
	MergeRec（按层）	**86.65***	**87.64***	**88.94***	79.21	74.91	91.43	**104.19***	87.27	79.87

表7：七种基线方法在四种骨干模型上的性能比较，即 RecFormer-base/large [13]和 BLaIR-base/large [7]。我们报告绝对 Recall@10 性能。除微调模型外，最佳结果以粗体标出，次佳结果以下划线标出。'*'表示 MergeRec 相较于最佳基线模型取得了统计显著提升（单尾 t 检验的 $p < 0.02$）。

骨干网络	方法	Avg.	Arts	美容	乐器	办公用品	食品储藏室	Sci.	体育	Toys
RecFormer-base	零样本	0.0767	0.1192	0.0445	0.0717	0.0850	0.0664	0.1309	0.0255	0.0706
	微调	0.1017	0.1567	0.0699	0.0994	0.1379	0.0895	0.1409	0.0358	0.0831
	联合学习	0.0815	0.1212	0.0556	0.0828	0.1012	0.0764	0.1127	0.0249	0.0773
	权重平均	0.0913	0.1429	0.0580	0.0903	0.1067	0.0838	0.1405	**0.0310**	0.0773
	任务算术	0.0904	0.1434	0.0569	0.0876	0.1156	0.0829	0.1394	0.0291	0.0685
	TIES	0.0926	0.1462	**0.0598**	0.0900	**0.1217**	0.0830	0.1377	0.0306	0.0717
	AdaMerging（按领域）	0.0802	0.1367	0.0410	0.0754	0.0893	0.0671	0.1305	0.0243	0.0774
	AdaMerging（按层）	0.0685	0.1053	0.0230	0.0670	0.0807	0.0588	0.1192	0.0233	0.0706
	MergeRec（按领域）	**0.0939***	**0.1507***	0.0585	0.0904	0.1162	**0.0855***	0.1420*	0.0301	**0.0776***
	MergeRec（按层）	0.0936*	0.1496*	0.0593	**0.0906***	0.1145	0.0849*	**0.1430***	0.0304	0.0766
RecFormer-large	零样本	0.0613	0.1138	0.0374	0.0470	0.0545	0.0415	0.1200	0.0205	0.0557
	微调	0.1038	0.1568	0.0721	0.1021	0.1388	0.0935	0.1463	0.0327	0.0877
	联合学习	0.0869	0.1304	0.0573	0.0815	0.1022	0.0844	0.1360	0.0299	0.0733
	权重平均	0.0947	0.1447	0.0628	0.0908	0.1098	**0.0902**	**0.1439**	**0.0323**	0.0827
	任务算术	0.0913	0.1432	0.0601	0.0901	0.1149	0.0861	0.1360	0.0287	0.0711
	TIES	0.0933	0.1458	0.0618	0.0922	**0.1224**	0.0869	0.1377	0.0291	0.0708
	AdaMerging（按领域）	0.0753	0.1307	0.0429	0.0792	0.0885	0.0574	0.1228	0.0190	0.0620
	AdaMerging（按层）	0.0735	0.1248	0.0377	0.0697	0.0842	0.0628	0.1216	0.0239	0.0631
	MergeRec（按领域）	**0.0965***	**0.1492***	**0.0650***	**0.0937***	0.1157	0.0900	0.1425	0.0310	**0.0847***
	MergeRec（按层）	0.0960*	0.1477*	0.0647*	0.0922	0.1141	0.0900	0.1432	0.0316	0.0843*
BLaIR-base	零样本	0.0409	0.0704	0.0224	0.0309	0.0376	0.0431	0.0759	0.0110	0.0355
	微调	0.0995	0.1535	0.0631	0.0980	0.1356	0.0913	0.1382	0.0320	0.0840
	联合学习	0.0832	0.1272	**0.0614**	**0.0819**	0.0991	0.0833	0.1034	**0.0297**	**0.0797**
	权重平均	0.0874	0.1412	0.0593	0.0768	0.1032	**0.0834**	**0.1381**	0.0286	0.0687
	任务算术	0.0612	0.0876	0.0337	0.0653	0.0835	0.0570	0.1013	0.0158	0.0452
	TIES	0.0825	0.1295	0.0492	0.0767	**0.1188**	0.0705	0.1223	0.0279	0.0650
	AdaMerging（按领域）	0.0603	0.1043	0.0329	0.0523	0.0608	0.0464	0.1008	0.0233	0.0614
	AdaMerging（按层）	0.0681	0.1122	0.0386	0.0675	0.0750	0.0525	0.1160	0.0238	0.0593
	MergeRec（按领域）	0.0869	**0.1449***	0.0564	0.0762	0.1104	0.0778	0.1320	0.0272	0.0706
	MergeRec（按层）	**0.0875**	0.1436*	0.0579	0.0774	0.1100	0.0802	0.1334	0.0280	0.0698
BLaIR-large	零样本	0.0333	0.0589	0.0205	0.0246	0.0257	0.0429	0.0501	0.0096	0.0339
	微调	0.0995	0.1534	0.0683	0.1003	0.1334	0.0923	0.1325	0.0346	0.0813
	联合学习	0.0842	0.1278	0.0553	0.0871	0.1048	0.0836	0.1040	0.0290	**0.0818**
	权重平均	0.0886	0.1423	0.0609	0.0841	0.1017	0.0860	0.1372	0.0299	0.0664
	任务算术	0.0785	0.1268	0.0540	0.0661	0.0993	0.0772	0.1240	0.0253	0.0552
	TIES	0.0905	0.1439	0.0622	0.0881	**0.1201**	0.0854	0.1303	**0.0324**	0.0613
	AdaMerging（按领域）	0.0689	0.1261	0.0426	0.0720	0.0683	0.0562	0.1039	0.0267	0.0556
	AdaMerging（逐层）	0.0778	0.1301	0.0504	0.0788	0.0833	0.0766	0.1114	0.0247	0.0675
	MergeRec（按领域）	0.0913*	0.1500*	0.0636*	**0.0921***	0.1074	0.0828	0.1363	0.0312	0.0673
	MergeRec（按层）	**0.0932***	**0.1515***	**0.0651***	0.0921*	0.1102	**0.0869***	**0.1394***	0.0312	0.0696

表8：在四种骨干模型（即 RecFormer-base/large [13]和 BLaIR-base/large [7]）上与七种基线方法的性能比较。我们报告 NDCG@10 的绝对性能。除微调模型外，最佳结果以粗体标出，次佳结果以下划线标出。"*"表示 MergeRec 相较于最佳基线模型取得了统计显著提升（单尾 t 检验的 $p < 0.02$）。

骨干网络	方法	Avg.	Arts	美容	乐器	办公用品	食品储藏室	Sci.	体育	Toys
RecFormer-base	零样本	0.0447	0.0717	0.0212	0.0415	0.0545	0.0381	0.0858	0.0118	0.0331
	微调	0.0662	0.1128	0.0344	0.0743	0.1036	0.0550	0.0924	0.0166	0.0407
	联合学习	0.0484	0.0760	0.0259	0.0566	0.0691	0.0455	0.0666	0.0116	**0.0356**
	权重平均	0.0551	0.0935	0.0263	0.0562	0.0728	0.0508	0.0921	**0.0137**	0.0351
	任务算术	0.0557	0.0975	0.0261	0.0565	0.0791	0.0505	0.0920	0.0130	0.0307
	TIES	0.0567	0.0983	**0.0274**	0.0569	**0.0842**	0.0506	0.0902	0.0134	0.0326
	AdaMerging（按领域）	0.0471	0.0851	0.0190	0.0448	0.0585	0.0391	0.0838	0.0113	0.0354
	AdaMerging（按层）	0.0408	0.0687	0.0110	0.0389	0.0532	0.0353	0.0763	0.0107	0.0322
	MergeRec（按领域）	**0.0570***	**0.0985**	0.0263	**0.0588***	0.0796	**0.0520***	0.0929*	0.0134	0.0346
	MergeRec（按层）	0.0565	0.0972	0.0266	0.0578*	0.0781	0.0516*	**0.0933***	0.0136	0.0341
RecFormer-large	零样本	0.0357	0.0720	0.0174	0.0260	0.0317	0.0235	0.0788	0.0096	0.0266
	微调	0.0667	0.1107	0.0346	0.0764	0.1015	0.0574	0.0954	0.0157	0.0419
	联合学习	0.0505	0.0798	0.0261	0.0508	0.0677	0.0499	0.0836	0.0134	0.0329
	权重平均	0.0573	0.0956	0.0283	0.0584	0.0757	0.0543	0.0941	**0.0143**	0.0372
	任务算术	0.0572	0.1036	0.0277	0.0593	0.0786	0.0526	0.0902	0.0128	0.0326
	TIES	0.0585	0.1042	0.0283	0.0616	**0.0833**	0.0530	0.0920	0.0131	0.0322
	AdaMerging（按领域）	0.0449	0.0853	0.0197	0.0473	0.0578	0.0319	0.0791	0.0088	0.0292
	AdaMerging（逐层）	0.0433	0.0786	0.0179	0.0410	0.0555	0.0357	0.0776	0.0110	0.0293
	MergeRec（按领域）	**0.0596***	**0.1066***	0.0293*	**0.0617**	0.0794	0.0540	0.0944	0.0138	0.0379*
	MergeRec（按层）	0.0591*	0.1034	**0.0295***	0.0598	0.0788	**0.0544**	**0.0950**	0.0141	**0.0380***
BLaIR-base	零样本	0.0218	0.0368	0.0103	0.0178	0.0216	0.0238	0.0426	0.0052	0.0165
	微调	0.0660	0.1119	0.0309	0.0742	0.1038	0.0582	0.0931	0.0153	0.0404
	联合学习	0.0518	0.0895	**0.0301**	**0.0588**	0.0702	**0.0512**	0.0633	**0.0141**	**0.0372**
	权重平均	0.0516	0.0879	0.0265	0.0497	0.0672	0.0498	**0.0881**	0.0126	0.0306
	任务算术	0.0365	0.0561	0.0160	0.0412	0.0543	0.0342	0.0620	0.0075	0.0209
	TIES	0.0493	0.0808	0.0226	0.0487	**0.0816**	0.0418	0.0768	0.0126	0.0296
	AdaMerging（按领域）	0.0332	0.0577	0.0152	0.0308	0.0367	0.0251	0.0615	0.0105	0.0279
	AdaMerging（按层）	0.0382	0.0637	0.0176	0.0398	0.0465	0.0288	0.0723	0.0106	0.0263
	MergeRec（按领域）	0.0517	**0.0910***	0.0253	0.0487	0.0732	0.0455	0.0858	0.0122	0.0316
	MergeRec（按层）	**0.0522**	0.0909*	0.0261	0.0497	0.0729	0.0477	0.0869	0.0125	0.0313
BLaIR-large	零样本	0.0172	0.0301	0.0091	0.0134	0.0141	0.0237	0.0278	0.0045	0.0150
	微调	0.0642	0.1093	0.0328	0.0743	0.0991	0.0565	0.0865	0.0160	0.0388
	联合学习	0.0524	0.0886	0.0265	**0.0635**	0.0752	**0.0521**	0.0621	0.0136	**0.0376**
	权重平均	0.0523	0.0910	0.0274	0.0533	0.0672	0.0513	0.0851	0.0135	0.0296
	任务算术	0.0468	0.0802	0.0254	0.0425	0.0656	0.0462	0.0775	0.0121	0.0253
	TIES	0.0550	0.0937	0.0289	0.0567	**0.0828**	0.0508	0.0847	**0.0145**	0.0279
	AdaMerging（按领域）	0.0386	0.0748	0.0191	0.0428	0.0421	0.0317	0.0623	0.0121	0.0242
	AdaMerging（按层）	0.0454	0.0810	0.0227	0.0497	0.0541	0.0456	0.0684	0.0113	0.0300
	MergeRec（按领域）	0.0542	0.0948*	0.0281	0.0584	0.0717	0.0494	0.0872*	0.0138	0.0300
	MergeRec（按层）	**0.0556***	**0.0958***	**0.0292***	0.0588	0.0743	0.0517	**0.0901***	0.0140	0.0310