arXiv:2302.03221v1 [cs.IR] 7 Feb 2023

基于外部注意力增强图卷积网络的轻量级跨域序列推荐

张金宇[†]、段慧川[†]、郭磊[†]、徐连成[†]和王新华[†⋆]

[†] 山东师范大学信息科学与工程学院，中国济南，250358
jinyuz1996@outlook.com, leiguo.cs@gmail.com,{hcduan, lcxu, wangxinhua}@sdnu.edu.cn

摘要。跨域序列推荐（CSR）是一项新兴但颇具挑战性的任务，它通过对多个领域中的交互进行建模，刻画重叠用户行为模式的演变。现有的 CSR 研究主要着眼于采用复合或深层结构来显著提升准确率，但这也给模型训练带来了沉重负担。此外，为了学习用户特定的序列表示，现有方法通常采用全局相关性加权策略（如自注意力机制），其计算复杂度呈二次增长。本文提出一种轻量级的外部注意力增强型图卷积网络框架，以解决上述挑战，称为 LEA-GCN。具体而言，我们仅保留邻域聚合组件并采用单层聚合协议（SLAP），使轻量级图卷积网络编码器能够更高效地捕捉两个领域中物品的协同过滤信号。为了进一步简化框架结构并聚合用户特定的序列模式，我们设计了一种新颖的双通道外部注意力（EA）组件，通过轻量级线性结构计算所有物品之间的相关性。我们在两个真实世界数据集上进行了大量实验，结果表明，与若干最先进的方法相比，LEA-GCN 在不影响准确率的前提下，模型体积更小、训练时间更短。

关键词：推荐系统 · 跨领域序列推荐 · 注意力机制 · 图神经网络 · 协同过滤

1 引言

序列推荐（SR）旨在从交互序列中捕捉用户的动态行为模式，近年来受到广泛关注，并已应用于许多领域，例如电子商务、在线检索和移动服务 [17,21,36,1,6]。尽管一些序列推荐方法

⋆ 通讯作者。

在许多热门任务中取得了巨大成功，但在从稀疏数据或冷启动场景中刻画用户偏好时仍面临挑战。因此，跨域序列推荐（CSR）正受到越来越多的关注，它通过利用其他领域的辅助信息来缓解上述问题 [37]。CSR 的核心思想是，向重叠用户推荐下一项；这类用户在同一时期内于多个领域均有可观察到的历史交互。

早期的跨领域序列推荐系统（CSR）采用基于循环神经网络（RNN）的结构 [26,31] 来捕捉混合序列中的序列依赖关系，但无法对跨领域实体之间的关联进行建模。随后，基于注意力的方法（例如 Chen 等人 [2] 和 Li 等人 [20], ）采用双重注意力结构，在领域之间有选择地传递用户的序列偏好，但难以挖掘序列转移内部的结构模式。近期提出的用于跨域序列推荐任务的基于图卷积网络（GCN）的方法 [37,10,12] 通过跨领域序列（CDS）图连接两个领域，并通过考虑结构信息传递细粒度的领域知识。然而，这类基于图的方法规模庞大、计算复杂度高，导致模型运行效率低下，阻碍了它们在内存有限的通用设备（例如图形处理器）上的部署。

为实现内存高效的轻量级推荐系统，近期研究 [4,23,28] 主要致力于压缩原始协同过滤矩阵，以提升推荐效率。然而，在嵌入转换过程中，原始表示之间的依赖关系可能会受到干扰，而且也无法充分建模用户在一段时间内的兴趣迁移。随后，为了从轻量化的角度探索用户偏好的演变与迁移，一些研究人员 [21,27] 将研究重点转向基于Transformer的模型。然而，Transformer [32] 中的自注意力机制具有二次复杂度，导致参数规模庞大。CSR中也存在类似问题，且情况更为严重，因为CSR需要在两个领域之间共享信息，从而使参数量翻倍。因此，开发轻量级CSR已成为持续发展的趋势，但同时也面临两项重大挑战：1）如前所述，当前的CSR开始采用图卷积网络 [37,10] 等复杂而深入的模型来学习用户的主要偏好，这给模型训练带来了很大负担，却对节点表示学习贡献甚微。2）全局相关性加权策略（例如，基于Transformer的方法） [12,21,27] 可能并非捕捉序列行为模式的最佳方案，因为它的计算复杂度极高，且忽略了双领域混合序列中物品的位置关系。

在这项工作中，我们提出了一种新颖的轻量级外部注意力增强图卷积网络（LEA-GCN），以应对上述挑战。具体而言，我们首先提取原始混合序列中每个物品的位置信息，以更好地捕捉重叠用户跨领域的行为演变。随后，我们构建跨领域序列（CDS）图，并

对用户和物品之间复杂的领域内关联进行建模，例如用户-项目交互以及各领域中物品的序列顺序。之后，我们通过两个步骤简化图卷积网络编码器：1）移除特征变换矩阵和非线性激活函数（即类似于 LightGCN [15] 对 NGCF [33]所做的处理），仅保留图卷积网络中的简单加权求和聚合器，以捕捉来自两个领域的协同过滤信号。2）随后，我们采用单层聚合协议（SLAP），以降低层传播的复杂度，同时避免 CDS 图中的高阶连通性造成的干扰。为应对第二项挑战，我们采用了一种新近提出的技术——外部注意力（EA）；该技术在计算机视觉（CV）领域受到了广泛关注 [7,13]。EA 使用两个外部记忆单元优化传统自注意力机制的计算复杂度，这与我们的轻量化目标不谋而合。然而，外部记忆单元独立于输入特征，因而难以对物品的协同过滤信号进行建模。为解决这一问题，我们设计了一种基于双通道 EA 的序列编码器，以学习用户特定的序列模式。它通过多头结构同时计算物品与外部记忆单元之间的相关性，并利用多层感知机（MLP）计算序列中每个物品的关系得分。本工作的主要贡献总结如下：

– 指出现有跨域序列推荐方法在参数规模和训练效率方面的缺陷后，我们提出了一种轻量级的基于图卷积网络的方案 LEA-GCN，用于实现内存高效的跨域序列推荐。– 我们通过简化网络结构并采用单层聚合协议来改进图卷积网络。随后，我们设计了双通道外部注意力，从轻量化的角度对用户的序列偏好进行建模。– 在两个真实世界数据集上进行的大量实验表明，与若干最先进的基线方法相比，LEA-GCN 性能更优，且所需参数更少。

2 相关工作

2.1 序列推荐

序列推荐系统（SRs）旨在对用户-物品交互序列进行建模 [21], ，已被证明能有效捕捉用户行为模式的演变 [17]。现有关于序列推荐的研究可分为传统方法和基于深度学习的方法。早期的传统序列推荐方法通常采用马尔可夫链假设来捕捉高阶序列模式 [14,5]。随着深度神经网络的发展，研究人员已将基于循环神经网络的方法 [30,34], 、基于图神经网络（GNN）的方法 [38,36,6],、基于 Transformer 的方法 [1,18,3], 以及自监督方法 [29,35] 应用于序列推荐任务。

这些方法具有强大的表征学习能力，但难以应对数据稀疏性或冷启动带来的挑战。

2.2 跨领域序列推荐

将其他领域的信息视为补充，跨领域序列推荐（CSR）方法可以缓解序列推荐（SR）中的数据稀疏性和冷启动问题 [25]。在早期探索中， π-net [26]和 PSJNet [31] 是两种用于 CSR 的基于RNN的解决方案，它们并行共享领域间的信息，同时学习两个领域的序列表示。随后，Zheng 等人 [37] 和 Guo等人 [10,12] 从基于图的视角研究 CSR，首先构建 CDS图，并通过注意力机制从局部和全局两个方面学习用户特定表示。另一个研究方向是基于注意力的方法，例如 Chen 等人 [2]和 Li 等人 [20], 通过双重注意力学习机制，将用户的序列偏好与候选项目进行匹配，从而提供跨领域推荐。随着神经网络深度的增加和模型结构日益复杂，这些 CSR 方法的参数规模或内存开销逐渐变得难以控制，超出了大多数传统设备的承载能力。

2.3 轻量级推荐

为了简化推荐系统的结构，使其更易于在各种设备上实现，轻量级推荐的概念受到了广泛关注 [28]。在传统方法中，近期研究主要聚焦于精简 DeepFM [24], DNN [23,22], 图卷积网络 [28],和 Transformer [27] 的结构，以提高推荐系统的内存效率。Li 等人 [21] 首次通过双重注意力网络为序列推荐（SR）引入轻量级方案，同时应对精简参数和从所有交互过的物品中发现时间信号这两项挑战。由于 CSR 通常需要辅助结构来连接多个领域，因此往往需要更多训练时间和更大的参数规模，但相关的轻量级方案大多尚未得到探索。

3 方法

3.1 预备知识

跨域序列推荐任务旨在通过对混合序列 [37] 中的历史交互进行建模，为重叠用户推荐下一个物品。假设 $U = \{U_1, U_2, \ldots, U_k, \ldots, U_p\}$ 是其历史行为在两个领域中均可获取的重叠用户集合，其中 $U_k \in \mathcal{U}$ （$1 \leq k \leq p$）表示领域 $\mathcal{U}$ 中的独立用户。令 S_H 为某个重叠用户的原始混合序列，我们进一步将 S_H 划分为 $S_A = \{A_1, A_2, \ldots, A_i, \ldots, A_m\}$ 和

$S_B = \{B_1, B_2, \ldots, B_j, \ldots, B_n\}$，它们分别表示领域 A 和领域 B 中的交互序列，其中 $A_i \in \mathcal{A}$ （$1 \leq i \leq m$）表示领域 A 中的物品，$B_j \in \mathcal{B}$ （$1 \leq j \leq n$）表示领域 B 中的物品。然后，我们令 $P_A = \{P_{A_1}, P_{A_2}, \ldots, P_{A_i}, \ldots, P_{A_m}\}$ 和 $S_B = \{P_{B_1}, P_{B_2}, \ldots, P_{B_j}, \ldots, P_{B_n}\}$ 分别表示序列 S_A 和 S_B 的位置信息，从而准确记录物品在原始混合序列 S_H 中的位置。

两个领域中所有候选项目被推荐的概率可表示为:

$$P(A_{i+1}|S_A, S_B) \sim f_A(S_A, S_B), \tag{1}$$
$$P(B_{j+1}|S_B, S_A) \sim f_B(S_B, S_A), \tag{2}$$

其中，$P(A_{i+1}|S_A, S_B)$ 表示基于 S_A 和 S_B，将 A_{i+1} 推荐为领域 A 中下一个消费项目的概率。$f_A(S_A, S_B)$ 表示用于估计该概率的学习函数。领域 B 的类似定义可表示为 $P(B_{j+1}|S_B, S_A)$ 和 $f_B(S_B, S_A)$。

3.2 概述

LEA-GCN 的核心思想是开发一种轻量级的基于图的方法来实现 CSR，同时不影响预测准确率。在 LEA-GCN 中，我们采用单层聚合协议（SLAP）优化 GCN 的结构，并利用外部注意力（EA）机制简化序列编码器。如图 3.2 所示，我们首先记录混合序列中所有物品的位置顺序，以保留用户跨领域行为模式的信息。其次，在保持混合序列中各领域内顺序不变的情况下，选取各领域的物品，从而得到领域 A 和领域 B 的子序列 S_A 和 S_B。随后，我们遵循与 DA-GCN [10] 相同的构成规则，构建 CDS图，以考虑用户与物品之间复杂的关联（即跨领域用户-项目交互关系和领域内项目-项目顺序关系）。构建 CDS图后，我们采用 Light-GCN [15] 图编码器，在 CDS图上线性传播嵌入，并采用 SLAP 学习节点表示，同时优化参数规模。随后，为了以较低的计算复杂度进一步捕捉用户交互中的序列模式，我们设计了一种基于双通道外部注意力（EA）的序列编码器。该编码器计算所有物品之间的相关性，并考虑从混合序列中提取的位置信息（即 $\boldsymbol{V}_A$ 和 $\boldsymbol{V}_B$）。由此得到的序列级表示可分别表示为 $\boldsymbol{H}_{S_A}$ 和 $\boldsymbol{H}_{S_B}$，对应两个领域。最后，我们将 $\boldsymbol{H}_{S_A}$ 和 $\boldsymbol{H}_{S_B}$ 拼接后输入预测层。

3.3 轻量级节点表示学习

图构建。受 Guo等人启发 [12], ，我们通过考虑两种关联来构建 CDS图，以连接两个领域：1）两个领域之间的用户-物品交互；2）两个领域内部的项目-项目序列转移

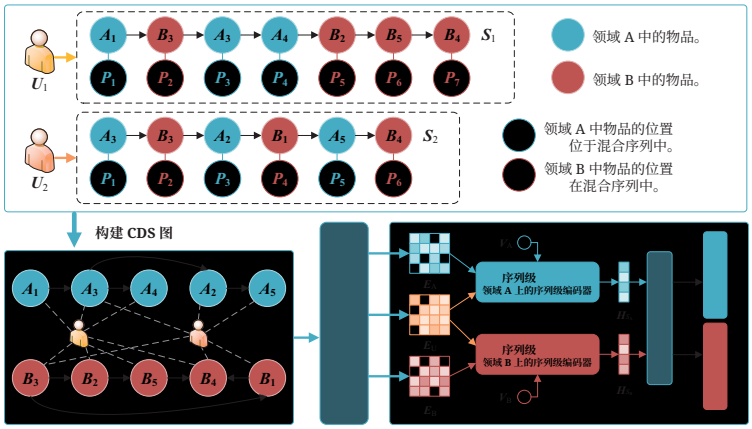

图 1. 我们提出的 LEA-GCN 概览，其中 S_1 和 S_2 分别表示两个不同重叠用户 U_1 和 U_2的混合序列。

，其中每个领域中的用户和物品都是节点，其关联关系则是边。然而，这种组合方法只保留了领域内的序列转移特征，却忽略了原始混合序列中物品的顺序依赖。为了解决上述缺陷，我们在构建 CDS图之前，额外记录混合序列中物品的位置信息，以便在序列级表示学习中使用。

然后，CDS图 G $\in \mathbb{R}^{(m+p+n)\times(m+p+n)}$ 可用矩阵形式表示为：

$$\mathcal{G} = \begin{bmatrix} \boldsymbol{A} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{B}, \end{bmatrix}; \qquad \boldsymbol{A} = \begin{bmatrix} \mathbf{0} & \boldsymbol{R}_{AU} \\ \boldsymbol{R}_{AU}^{T} & \mathbf{0}, \end{bmatrix}; \qquad \boldsymbol{B} = \begin{bmatrix} \mathbf{0} & \boldsymbol{R}_{BU} \\ \boldsymbol{R}_{BU}^{T} & \mathbf{0}, \end{bmatrix}, \tag{3}$$

其中，$\boldsymbol{A} \in \mathbb{R}^{(m+p)\times(m+p)}$ 和 $\boldsymbol{B} \in \mathbb{R}^{(p+n)\times(p+n)}$ 分别表示两个领域的压缩拉普拉斯矩阵，$\boldsymbol{R}_{AU} \in \mathbb{R}^{m\times p}$ 和 $\boldsymbol{R}_{BU} \in \mathbb{R}^{p\times n}$分别表示领域 A 和领域 B 的用户-物品交互矩阵，$\boldsymbol{R}_{AU}^{T} \in \mathbb{R}^{m\times p}$ 和 $\boldsymbol{R}_{BU}^{T} \in \mathbb{R}^{p\times n}$ 为其转置矩阵。

单层聚合协议（SLAP）。早期提出的先进方法 NGCF [33] 和 LightGCN [15]，采用多层聚合协议（MLAP），逐层传播节点嵌入，已被证明有效。MLAP 的核心思想是利用用户-物品二部图上的高阶连通性，获取用户与物品之间潜在的关联，从而提升节点表示学习的性能。第 l 层的节点表示学习可详细表示为：

$$\boldsymbol{E}^{(l)} = H(\boldsymbol{E}^{(l-1)}, \mathcal{G}), \tag{4}$$

其中， $\mathcal{G}$ 是用户-物品二部图， $\boldsymbol{E}^{(l)}$ 表示第 l 层的节点表示， $\boldsymbol{E}^{(l-1)}$ 表示上一层的节点表示。 H（·）表示邻居聚合函数。

然而，MLAP 可能并不适用于 CDS图。例如，在学习领域 A 的物品表示时，若考虑领域 B 中物品的高阶连接性，也会引入领域 B 的高阶领域特定结构信息，而这些信息可能成为领域 A 的噪声，干扰重叠用户的兴趣表达。因此，我们提出一种单层传播规则（即单层聚合协议（SLAP）），以便在 CDS图上进行一阶邻居聚合：

$$\boldsymbol{E}_U = \sum_{k \in U} \left(\sum_{i \in S_A} \frac{1}{\sqrt{|N_U|}\sqrt{|N_A|}} \boldsymbol{e}_{\boldsymbol{i} \to \boldsymbol{k}} + \sum_{j \in S_B} \frac{1}{\sqrt{|N_U|}\sqrt{|N_B|}} \boldsymbol{e}_{\boldsymbol{j} \to \boldsymbol{k}} \right); \tag{5}$$

$$\boldsymbol{E}_A = \sum_{i \in S_A} \sum_{k \in U} \frac{1}{\sqrt{|N_U|}\sqrt{|N_A|}} \boldsymbol{e}_{\boldsymbol{k} \to \boldsymbol{i}}; \tag{6}$$

$$\boldsymbol{E}_B = \sum_{j \in S_B} \sum_{k \in U} \frac{1}{\sqrt{|N_U|}\sqrt{|N_B|}} \boldsymbol{e}_{\boldsymbol{k} \to \boldsymbol{j}}, \tag{7}$$

其中， $\boldsymbol{e}_{\boldsymbol{i} \to \boldsymbol{k}}$ 和 $\boldsymbol{e}_{\boldsymbol{j} \to \boldsymbol{k}}$ 表示从物品传递给重叠用户的消息， $\boldsymbol{e}_{\boldsymbol{k} \to \boldsymbol{i}}$ 和 $\boldsymbol{e}_{\boldsymbol{k} \to \boldsymbol{j}}$ 分别表示从用户传递到领域 A 和领域 B 中物品的消息， N_U 表示与物品 A_i 或物品 B_j交互过的重叠用户集合， N_A 和 N_B 表示与用户 u交互过的物品集合。采用 SLAP 后，领域间的知识能够以较少受到领域特定信息影响的方式进行传递，同时还能缩减参数规模。

为支持我们的观点，我们开展了一系列消融实验，以探究层深度的影响。由于篇幅限制，我们仅在表 1 中报告 DOUBAN（即一个真实世界数据集，将在第 4.1 节中详细介绍）的实验结果。

如表 1 所示，我们比较了 LEA-GCN 在不同层数下的表现与两种最先进的基于图卷积网络的方法（即 LightGCN [15]和 DA-GCN [10]）。为了使 LightGCN 能够与另外两种跨域序列推荐方法进行比较，我们同时报告了其在两个领域上的表现。由此得到以下观察结果：1）增加层数可以提升 LightGCN 的表现，表明在单一领域内的用户—物品图上传播嵌入时，考虑高阶连通性是有效的。2）采用 SLAP 的 DA-GCN 和 LEA-GCN 优于采用 MLAP 的版本，这支持了我们的假设：在 CDS图 上简单地聚合高阶连通性中的消息，可能会给节点表示学习带来更多噪声；同时也证明了在跨域序列推荐任务中采用 SLAP 的重要性。

表 1. 不同层数下 LightGCN、DA-GCN 和我们提出的基于图卷积网络的解决方案的性能（%）对比。

领域		A			B		
层数	方法	RC10	MRR10	NDCG10	RC10	MRR10	NDCG10
@ 1 层	LightGCN	78.03	75.06	48.24	68.28	52.03	37.52
	DA-GCN	83.55	80.84	51.53	71.91	58.31	40.67
	LEA-GCN	83.83	81.22	**52.06**	76.14	66.15	**45.01**
@ 2 层	LightGCN	78.33	75.25	48.89	68.19	52.00	37.66
	DA-GCN	83.12	80.55	50.92	71.42	57.45	40.16
	LEA-GCN	83.53	81.08	51.35	74.63	63.67	42.91
@ 3 层	LightGCN	78.52	75.35	48.93	68.32	52.11	37.79
	DA-GCN	81.99	80.08	49.35	71.08	57.17	39.85
	LEA-GCN	83.40	81.02	51.05	74.22	63.29	42.11

3.4 基于外部注意力（EA）的序列编码器

自注意力通过组合自身的值来优化输入序列表示，它只考虑序列内物品之间的关系，却忽略了不同序列中物品之间隐含的关系 [13]。此外，O(N^2) 的高计算复杂度也是使用自注意力的一大缺点。为了准确衡量不同物品对序列的贡献，同时降低计算开销，我们以外部注意力（EA）机制取代基于自注意力的算法 [12,27] 。该机制已在图像分类、目标检测和语义分割任务中取得了巨大成功 [13]。

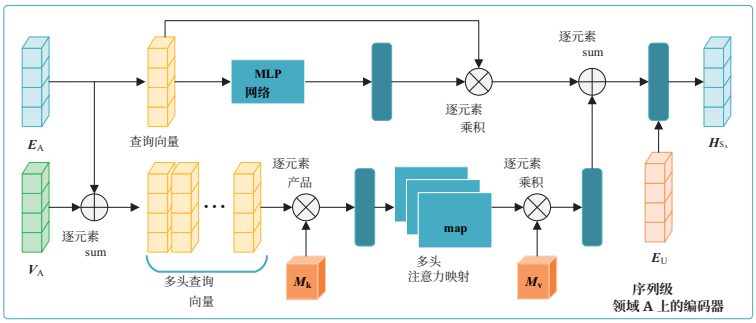

图 2. 我们提出的双通道外部注意力序列编码器的工作流程（以领域 A 为例），其中 M_k 和 M_v 是两个不同的外部记忆单元。

与自注意力机制不同，外部注意力（EA）使用外部存储单元来保存全局共享权重。如图3.4所示，我们设计了一种双通道复合线性结构来实现外部注意力（以领域 A 为例）。具体而言，在第一个通道中，我们将混合序列的位置信息 $\boldsymbol{V}_A$ 附加到物品表示$\boldsymbol{E}_A$上。随后，我们计算物品与两个外部记忆单元（即 $\boldsymbol{M}_k$ 和 $\boldsymbol{M}_v$）之间的外部注意力，它们分别充当键矩阵和值矩阵。分数计算可表示为:

$$\boldsymbol{T}_{S_A}^{(1)} = Norm[(\boldsymbol{E}_A + \alpha \boldsymbol{V}_A)\boldsymbol{M}_k^T]; \tag{8}$$

$$\boldsymbol{H}_{S_A}^{(1)} = \boldsymbol{T}_{S_A}^{(1)} \boldsymbol{M}_v^T, \tag{9}$$

其中， α 是用于控制位置信息参与程度的超参数， $\boldsymbol{T}_{S_A}^{(1)}$ 是外部注意力图, $\boldsymbol{H}_{S_A}^{(1)}$ 表示第一个通道中序列 S_A 的最终输出嵌入。对于 S_A中的第 i 个物品 A_i , 计算过程可进一步详细表示为:

$$\boldsymbol{a}_{i,j}^{(1)} = Norm[(e_{A_i} \oplus (\alpha \cdot v_{A_i}))m_j^k]; \tag{10}$$

$$\boldsymbol{H}_{S_A}^{(1)} = \sum_{i=1}^{|S_A|} \boldsymbol{a}_{i,j}^{(1)} m_i^v, \tag{11}$$

其中， $\boldsymbol{a}_i^{(1)}$, $_j$ 是第 i 个物品 e_{A_i} 与矩阵 M_k的第 j 行 m_j^k之间的成对亲和度。 v_{A_i} 表示 A_i 的位置信息，而 m_i^v 表示第二个记忆单元的第 i 行。

受 Transformer [32], 的启发，我们采用多头方式 [13] 来提升外部注意力的能力，从而更好地捕捉物品之间的不同关系:

$$Z_h^{(1)} = ExternalAttention(\boldsymbol{E}_A, \boldsymbol{V}_A, \boldsymbol{M}_k, \boldsymbol{M}_v); \tag{12}$$

$$\boldsymbol{H}_{S_A}^{(1)} = Concat(Z_h^{(1)}, \ldots, Z_\beta^{(1)})\boldsymbol{W}_1, \tag{13}$$

其中， $\boldsymbol{W}_1$ 是用于对齐输入和输出维度的线性变换矩阵， Z_h 是第 h 个注意力头， β 控制注意力头的数量。在实验中，我们在 [1, 2, 4, 8, 16] 中搜索 β ，并在第4.5节报告实验结果。

尽管外部注意力可以很好地解决计算复杂度问题，但其最终的输出表示由 M_v生成。由于外部记忆单元与输入特征（即 E_A）相互独立，因此在建模物品的协同过滤信号方面存在不足。为解决上述问题，在第二个通道中，使用带有平滑归一化层的多层感知机网络（受Guo等人 [12]的启发）来提取 E_A的协同过滤信号。我们测量 e_i 与 e_j 之间的相关性如下:

$$\boldsymbol{a}_{i,j}^{(2)} = Norm(f(e_{A_i}, e_{A_j})) = Norm(\boldsymbol{W}_3^T \boldsymbol{ReLU}(\boldsymbol{W}_2[e_{A_i} \oplus e_{A_j}] + \boldsymbol{b})), \tag{14}$$

其中， $f(\cdot)$ 是由多层感知机网络实现的评分函数， $\boldsymbol{W}_2$ 和 $\boldsymbol{W}_3$是两个权重矩阵， $\boldsymbol{b}$ 是偏置向量。然后，我们可以从第二个通道中获得序列表示 $\boldsymbol{H}_{S_A}^{(2)}$：

$$\boldsymbol{H}_{S_A}^{(2)} = \sum_{i=1}^{|S_A|} \boldsymbol{a}_{i,j}^{(2)} e_{A_i}. \tag{15}$$

因此，领域 A 的用户特定序列级表示 $\boldsymbol{H}_{S_A}$ 可表示为：

$$\boldsymbol{H}_{S_A} = Concat((\boldsymbol{H}_{S_A}^{(1)} + \boldsymbol{H}_{S_A}^{(2)}), \boldsymbol{E_U}). \tag{16}$$

3.5 预测层

完成序列表示学习后，LEA-GCN 分别获得领域 A 和领域 B 的序列嵌入 $\boldsymbol{H}_{S_A}$ 和 $\boldsymbol{H}_{S_B}$。随后，为了利用两个领域的信息，我们将它们拼接后输入预测层：

$$P(A_{i+1}|S_A, S_B) = softmax(\boldsymbol{W}_A \cdot [\boldsymbol{H}_{S_A}, \boldsymbol{H}_{S_B}]^{\mathrm{T}} + \boldsymbol{b}_A); \tag{17}$$

$$P(B_{j+1}|S_B, S_A) = softmax(\boldsymbol{W}_B \cdot [\boldsymbol{H}_{S_B}, \boldsymbol{H}_{S_A}]^{\mathrm{T}} + \boldsymbol{b}_B), \tag{18}$$

其中， $\boldsymbol{W}_A$ 和 $\boldsymbol{W}_B$ 分别是领域 A 和领域 B 中所有物品的权重矩阵； $\boldsymbol{b}_A$ 和 $\boldsymbol{b_A}$ 是两个领域的偏置项。然后，为避免 π-net [26] 和 DA-GCN [10], 出现跷跷板现象，我们采用交叉熵损失，并在两个领域上分别独立优化：

$$\mathcal{L}_A = -\frac{1}{|\mathcal{S}|} \sum_{S_A, S_B \in \mathcal{S}} \sum_{A_i \in S_A} \log P(A_{i+1}|S_A, S_B), \tag{19}$$

$$\mathcal{L}_B = -\frac{1}{|\mathcal{S}|} \sum_{S_B, S_A \in \mathcal{S}} \sum_{B_j \in S_B} \log P(B_{j+1}|S_B, S_A), \tag{20}$$

其中， $\mathcal{S}$ 表示两个领域中的训练序列。

4 实验

我们在两个真实世界数据集上进行了大量实验，以验证 LEA-GCN 的有效性。在本节中，我们旨在回答以下研究问题（RQ）：

RQ1：LEA-GCN 能否减轻模型的参数负担？LEA-GCN 的训练效率如何？RQ2：与其他最先进的基线方法相比，LEA-GCN 的表现如何？我们的轻量化策略是否会导致推荐性能下降？RQ3：在混合输入序列中考虑物品的位置关系是否有帮助？通过外部注意力为两个领域学习序列表示是否有效？RQ4：超参数如何影响 LEA-GCN 的性能？

4.1 实验设置

数据集与评估协议。我们在两个真实世界数据集（即 DOUBAN [39] 和 AMAZON [8]）上评估 LEA-GCN。DOUBAN 包含领域 A 和领域 B 中重叠用户的历史交互（即豆瓣电影和豆瓣图书），这些数据收集自知名的中国社交媒体平台豆瓣[1] [39]。AMAZON 是 Fu等人收集的产品评论数据集 [8]。它包含用户在两个不同亚马逊[2] 平台上的评论行为，即亚马逊图书（领域 A）和亚马逊电影（领域 B）。如表 2 所示，我们从两个数据集的全部混合序列中随机选取 80% 作为训练集，其余 20% 作为测试集。此外，在对两个数据集进行预处理时，我们会过滤掉历史交互少于十次的冷用户，以及出现次数少于五次的冷物品 [17]。

在评估时，我们首先将每个混合序列中最后两个观测到的物品作为两个领域的真实物品。其次，我们采用三种常用指标（即 RC@10、MRR@10 和 NDCG@10） [10,11] 在测试集上评估每个实例，并报告其平均值。

表 2. 数据集统计信息，其中 A 和 B 表示两个数据集中的不同领域。

数据集	豆瓣		亚马逊	
领域	**A**	**B**	**A**	**B**
物品	14,636	2,940	126,526	61,362
交互	607,523 360	,798 1	,678,006 978	,226
用户	6,582		9,204	
序列（训练集）	42,062		90,574	
序列（测试集）	10,431		14,463	

基线方法。为验证 LEA-GCN 的性能，我们将所提出的方法与以下基线进行比较：1）传统推荐：NCF [16], NGCF [33], 和 LightGCN [15]。我们将传统方法适配为使用序列输入，并报告它们在各个领域的实验结果。2）序列推荐：GRU4REC [17] 和 HRNN [30]。我们报告它们在各个领域的性能。3）跨域序列推荐： π- Net [26], PSJNet [31], DA-GCN [10], 和 TiDA-GCN [12]。

[1] http://www.douban.com/ [2] http://jmcauley.ucsd.edu/data/amazon/

实现细节。我们使用 TensorFlow 实现 LEA-GCN[3]，并通过 NVIDIA Tesla K80M GPU 加速模型训练。对于参数，我们采用 Xavier [9] 进行初始化，并使用 Adam [19]进行优化。训练模型时，我们将批次大小设为 256，丢弃率设为 0.1，领域 A 和领域 B 的学习率分别设为 0.002 和 0.004。对于 LEA-GCN，我们将嵌入维度设为 16，正则化比率设为 1e-7。我们在 [0-1] 范围内以 0.1 为步长搜索超参数 α，以调整位置信息的参与程度；并在 [1, 2, 4, 8, 16] 范围内探索注意力头数量 β，以使 LEA-GCN 达到最佳性能。我们在 4.4 节详细介绍 α 和 β 的实验结果。此外，为了使结果具有可比性，我们将所有参考基线的嵌入维度统一设为 16。对于其他超参数，我们参考各论文中的最佳设置，并在两个数据集上进行微调。

4.2 参数规模与训练效率（RQ1）

在本节中，我们首先通过调整 [0.2 - 1.0] 中输入数据的比例，在豆瓣和亚马逊上进行一系列实验，以测量模型训练所消耗的时间。其次，我们通过将 LEA-GCN 的参数规模与两个最具竞争力的基线方法（即 PSJNet 和 TiDA-GCN）进行比较，分析其轻量化建模能力。由此得到以下观察结果：1）从图 3（a）和（b）可以看出，LEA-GCN 的训练耗时低于 TiDA-GCN 和 PSJNet，这表明 LEA-GCN 具有更高的训练效率，并且能够扩展至大规模数据集。2）从图 3（c）和（d）可以看出，LEA-GCN 和 LEA-All 所需的参数远少于 PSJNet 和 TiDA-GCN，从而对 RQ1 给出了肯定回答。需要注意的是，LEA-All 是一种变体方法，仅保留 LEA-GCN 中的图卷积网络编码器。

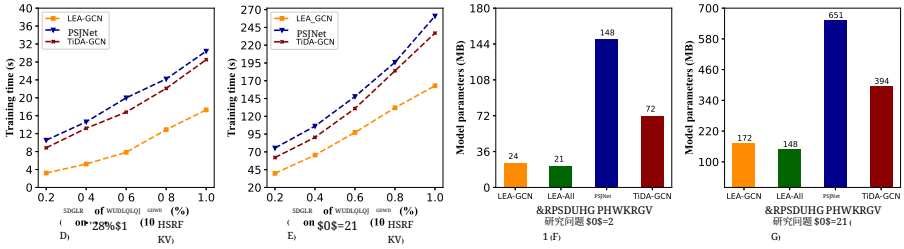

图3. LEA-GCN 与 PSJNet 和 TiDA-GCN 的时间消耗和参数规模对比。

[3] https://github.com/JinyuZ1996/LEA-GCN

4.3 推荐准确率表现（RQ2）

表 3 展示了 LEA-GCN 与其他最先进方法在两个数据集上的实验结果。观察结果总结如下：1）LEA-GCN 在亚马逊的两个领域均取得最佳性能，并且在豆瓣的大多数评估指标上优于其他最先进的基线方法，这表明 LEA-GCN 中的轻量化策略对预测准确率影响甚微，甚至能够提升模型性能。2）跨域序列推荐解决方案优于其他最先进方法（即传统推荐算法和序列推荐算法），表明同时建模用户的序列偏好和跨领域特征十分重要。3）LEA-GCN 优于所有跨域序列推荐基线方法，说明通过轻量化结构从 CDS图中建模用户特定偏好的方法有效，同时也体现了外部注意力在捕获用户序列模式方面的优势。

表 3. 各方法在豆瓣和亚马逊上的实验结果（%）。请注意，粗体数值表示相应指标的最佳结果。显著提升以† 标记（配对样本 t 检验， $p < .05$）。

数据集	豆瓣						亚马逊					
领域	**A**			**B**			**A**			**B**		
指标 (@10)	RC	平均倒数排名	归一化折损累计增益	RC	平均倒数排名	归一化折损累计增益	RC	平均倒数排名	归一化折损累计增益	RC	平均倒数排名	归一化折损累计增益
NCF	69.75	58.05	40.26	35.24	23.29	18.28	15.59	11.30	6.12	17.38	13.20	5.29
NGCF	79.21	77.82	49.12	67.37	54.41	37.26	21.52	18.74	12.08	26.55	25.37	17.45
LightGCN	78.52	75.35	48.93	68.32	52.11	37.79	21.67	17.41	10.13	26.49	24.23	15.09
GRU4REC	80.13	75.41	47.81	66.66	54.10	38.03	21.51	17.11	9.66	24.51	22.13	12.94
HRNN	81.25	77.90	48.88	68.32	54.99	38.93	21.92	17.30	9.87	25.10	22.48	13.44
π-net	83.22	80.71	51.22	69.54	55.72	39.18	24.33	20.52	11.80	27.66	25.03	16.20
PSJNet	83.54	80.96	51.72	71.59	58.36	40.71	25.03	21.09	13.54	31.24	28.35	18.93
DA-GCN	83.55	80.84	51.53	71.91	58.31	40.67	24.62	20.91	13.18	31.12	28.21	18.85
TiDA-GCN	83.68	81.27	52.02	72.56	60.27	41.38	25.05	21.23	14.68	32.84	29.65	19.12
LEA-GCN	**83.83**†	81.22	52.06	**76.14**†	**66.15**†	**45.01**†	**25.47**†	**21.57**†	**14.93**†	**33.97**†	**30.65**†	**20.46**†

4.4 消融研究 (RQ3)

本节中，我们在豆瓣和亚马逊上进行了一系列消融研究，以探究不同组件对 LEA-GCN 的影响。受篇幅限制，我们仅报告豆瓣上的结果。如表4所示，LEA-Pos 是一种变体，它禁用了混合序列中位置信息的参与。LEA-EA 是另一种变体模型，它移除了基于外部注意力的序列编码器。LEA-All 是一种同时禁用位置信息和外部注意力的变体。表4中的观察结果总结如下：1）LEA-GCN 的表现优于 LEA-All 和 LEA-Pos，说明原始混合序列中物品的位置信息对于学习重叠用户的序列特征至关重要。

表4. DOUBAN数据集上的消融研究实验结果（%）。

领域	A			B		
指标(@10)	RC	平均倒数排名	归一化折损累计增益	RC	平均倒数排名	归一化折损累计增益
LEA-Pos	83.67	81.16	52.00	75.10	64.35	44.38
LEA-EA	82.62	79.14	50.43	72.82	61.92	42.41
LEA-All	82.12	78.10	49.64	71.28	59.43	41.28
LEA-GCN	83.83	81.22	52.06	76.14	66.15	45.01

2）LEA-GCN 的表现优于 LEA-EA，说明外部注意力组件有助于序列级表示学习。

4.5 超参数分析（RQ4）

超参数 α 控制附加到两个领域的物品表示中的位置信息所占的比重。图4(a)至(f)展示了 LEA-GCN 在不同 $\alpha \in [0 - 1]$下的性能。实验结果证明，利用原始混合序列中的位置信息十分重要。不过，不宜将其视为与序列表示同等重要。

超参数 β 控制多头外部注意力组件的注意力头数量。图4(g)至(l)中的实验结果表明，只有注意力头数量适当时，外部注意力机制才有助于序列表示学习过程。

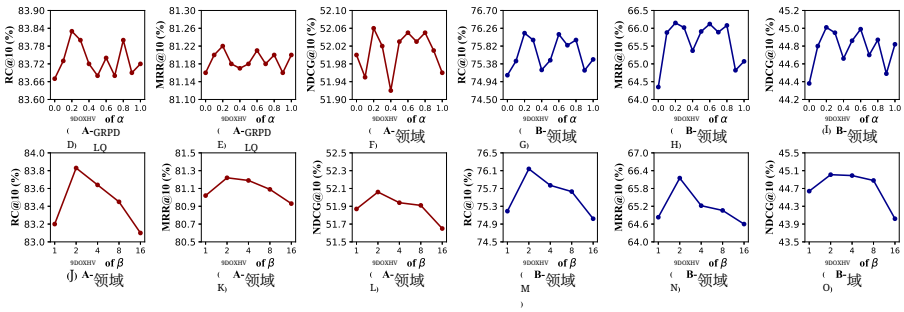

图4. 超参数 α 和 β 对豆瓣的影响。

5 结论

在这项工作中，我们提出了一种用于 CSR 的轻量级基于图卷积网络的解决方案，该方案同时简化了图卷积网络的结构并优化了计算

复杂度。具体而言，我们仅保留邻域聚合，以缩小图卷积网络编码器的参数规模，并提出单层聚合协议（SLAP），用于在 CDS图上传播嵌入。随后，我们设计了一种基于双通道外部注意力（EA）的序列编码器，通过更轻量的线性结构计算所有物品之间的相关性。在两个真实世界数据集上的实验结果证明了我们轻量级解决方案的优越性。

参考文献

1. Ai, Z., Wang, S., Jia, S., Guo, S.: 面向序列推荐的核心兴趣聚焦自注意力。载于：DASFAA。第 306–314 页（2022） 2. Chen, C., Guo, J., Song, B.: 多维集成的基于会话推荐中的双重注意力迁移。载于：SIGIR。第 869–878 页（2021） 3. Chen, H., Lin, Y., Pan, M., Wang, L., Yeh, C.M., Li, X., Zheng, Y., Wang, F., Yang, H.: 去噪自注意力序列推荐。载于：RecSys。第 92–101 页（2022） 4. Chen, T., Yin, H., Zheng, Y., Huang, Z., Wang, Y., Wang, M.: 学习弹性嵌入，以定制端侧推荐系统。载于：KDD。第 138–147 页（2021） 5. Cheng, C., Yang, H., Lyu, M.R., King, I.: 下一站，你想去哪里：连续兴趣点推荐。载于：IJCAI。第 2605–2611 页（2013） 6. Dong, X., Jin, B., Zhuo, W., Li, B., Xue, T.: Sirius：基于特征增强图神经网络的序列推荐。载于：DASFAA。第 315–320 页（2021） 7. Fang, J., Yang, C., Shi, Y., Wang, N., Zhao, Y.: 用于裂缝检测的基于外部注意力的 TransUNet 和标签扩展策略。TITS，第 19054–19063 页（2022） 8. Fu, W., Peng, Z., Wang, S., Xu, Y., Li, J.: 在跨域推荐系统中深度融合评论与内容，为冷启动用户提供推荐。载于：AAAI。第 94–101 页（2019） 9. Glorot, X., Bengio, Y.: 理解深度前馈神经网络的训练难点。载于：AISTATS。第 249–256 页。JMLR（2010） 10. Guo, L., Tang, L., Chen, T., Zhu, L., Nguyen, Q.V.H., Yin, H.: DA-GCN：面向共享账户跨域序列推荐的领域感知注意力图卷积网络。载于：IJCAI。第 2483–2489 页（2021） 11. Guo, L., Zhang, J., Chen, T., Wang, X., Yin, H.: 强化学习增强的共享账户跨域序列推荐。TKDE（2022） 12. Guo, L., Zhang, J., Tang, L., Chen, T., Zhu, L., Yin, H.: 面向共享账户跨域序列推荐的时间间隔增强图神经网络。TNNLs（2022） 13. Guo, M.H., Liu, Z.N., Mu, T.J., Hu, S.M.: 超越自注意力：使用两个线性层实现视觉任务的外部注意力。TPAMI（2022） 14. He, R., Kang, W., McAuley, J.J.: 基于翻译的推荐。载于：RecSys。第 161–169 页（2017） 15. 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） 16. 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） 17. Hidasi, B., Karatzoglou, A., Baltrunas, L., Tikk, D.: 基于循环神经网络的会话推荐。载于：ICLR（2016）

18. Kang, W., McAuley, J.J.: 自注意力序列推荐。载于：ICDM。第 197–206 页（2018） 19. Kingma, D.P., Ba, J.: Adam：一种随机优化方法。载于：ICLR （ 2015 ） 20. Li, P., Jiang, Z., Que, M., Hu, Y., Tuzhilin, A.: 面向跨域点击率预测的双重注意力序列学习。载于：KDD。第 3172–3180 页（2021） 21. Li, Y., Chen, T., Zhang, P., Yin, H.: 轻量级自注意力序列推荐。载于：CIKM。第 967–977 页（2021） 22. Lian, D., Wang, H., Liu, Z., Lian, J., Chen, E., Xie, X.: Lightrec：内存与搜索高效的推荐系统。载于：WWW。第 695–705 页（2020） 23. Liu, H., Zhao, X., Wang, C., Liu, X., Tang, J.: 深度推荐系统中的自动嵌入维度搜索。载于：SIGIR。第 2307–2316 页（2020） 24. Liu, S., Gao, C., Chen, Y., Jin, D., Li, Y.: 推荐系统的可学习嵌入维度。载于：ICLR（2021） 25. Ma, M., Ren, P., Chen, Z., Ren, Z., Zhao, L., Liu, P., Ma, J., de Rijke, M.: 面向跨域序列推荐的混合信息流。TKDD，第 64:1–64:32 页（2022） 26. Ma, M., Ren, P., Lin, Y., Chen, Z., Ma, J., Rijke, M.d.： π-net：面向共享账户跨域序列推荐的并行信息共享网络。载于：SIGIR。第 685–694 页（2019） 27. Mei, M.J., Zuber, C., Khazaeni, Y.: 用于下一物品商品推荐的轻量级Transformer。载于：RecSys。第 546–549 页（2022） 28. Miao, H., Li, A., Yang, B.: 元路径增强的轻量级图神经网络，用于社交推荐。载于：DASFAA。第 134–149 页（2022） 29. Qiu, R., Huang, Z., Yin, H., Wang, Z.: 针对序列推荐中表征退化问题的对比学习。载于：WSDM。第 813–823 页（2022） 30. Quadrana, M., Karatzoglou, A., Hidasi, B., Cremonesi, P.: 使用分层循环神经网络实现个性化基于会话的推荐。载于：RecSys。第 130–137 页（2017） 31. Sun, W., Ma, M., Ren, P., Lin, Y., Chen, Z., Ren, Z., Ma, J., De Rijke, M.: 面向共享账户跨域序列推荐的并行分拆合并网络。TKDE（2021） 32. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention Is All You Need。载于：NIPS。第 5998–6008 页（2017） 33. Wang, X., He, X., Wang, M., Feng, F., Chua, T.: 神经图协同过滤。载于：SIGIR。第 165–174 页（2019） 34. Wu, C., Ahmed, A., Beutel, A., Smola, A.J., Jing, H.: 循环推荐网络。载于：WSDM。第 495–503 页（2017） 35. 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） 36. Zang, Y., Liu, Y., Chen, W., Li, B., Li, A., Yue, L., Ma, W.: GISDCN：基于图的插值序列推荐模型，采用可变形卷积网络。载于：DASFAA。第 289–297 页（2022） 37. Zheng, X., Su, J., Liu, W., Chen, C.: DDGHM：采用混合度量训练的双动态图，用于跨域序列推荐。载于：MM。第 471–481 页（2022） 38. Zheng, Y., Liu, S., Li, Z., Wu, S.: DGTN：用于基于会话推荐的双通道图转移网络。载于：ICDM。第 236–242 页（2020） 39. Zhuang, F., Zhou, Y., Ying, H., Zhang, F., Ao, X., Xie, X., He, Q., Xiong, H.: 通过跨域新颖性寻求特征挖掘实现序列推荐。JCST，第 305–319 页（2020）