计算机研究与发展
Journal of Computer Research and Development
DOI: 10.7544/issn1000-1239.202440197
61(8): 1930−1944, 2024

基于三分支图外部注意力网络的轻量化跨域序列推荐

张劲羽　马晨曦　李　超　赵中英
（山东科技大学计算机科学与工程学院　山东青岛　266590）
（jinyuz1996@outlook.com）

Towards Lightweight Cross-Domain Sequential Recommendation via Tri-Branches Graph External Attention Network

Zhang Jinyu, Ma Chenxi, Li Chao, and Zhao Zhongying
（*College of Computer Science and Engineering*, *Shandong University of Science and Technology*, *Qingdao*, *Shandong* 266590）

Abstract　Cross-domain sequential recommendation (CSR) aims to capture the behavioral preferences of users by modeling their historical interaction sequences in multiple domains, thus providing personalized cross-domain recommendations. Recently, researchers have started integrating graph convolution networks (GCNs) into CSR to model complicated associations among users and items. However, due to their complicated structure, most graph-based CSR methods are usually accompanied by high computational complexity or memory overhead, making them difficult to deploy on resource-constrained edge devices. Besides, existing lightweight graph-based CSR methods tend to employ single layer aggregating protocol (SLAP) to propagate embeddings on cross-domain sequential graphs (CSG). Such a strategy indeed aids the GCNs in circumventing cross-domain noise interference caused by high-order neighborhood aggregation strategies. However, it also shields GCN from mining high-order sequential relationships within individual domains. To this end, we introduce a lightweight tri-branches graph external attention network (TEA-Net). Specifically, we separate the original CSG into three parts including two inner-domain sequential graphs and an inter-domain sequential graph and devise a parallel tri-branches graph convolution network to learn the node representations. This structure can simultaneously consider the first-order inter-domain correlations and the high-order inner-domain connectivity without introducing additional cross-domain noises. Besides, we propose an improved external attention (EA) component without the nonlinear channel, which captures the sequential dependency among items at a lower cost and shares attention weights across multiple branches. We conduct extensive experiments on two large-scale real-world datasets to verify the performance of TEA-Net. The experimental results demonstrate the superiority of TEA-Net in both the lightweight performance and the prediction accuracy compared with several state-of-the-art methods.

Key words　recommender systems; lightweight cross-domain sequential recommendation; external attention; graph neural network; collaborative filtering

摘　要　跨域序列推荐（cross-domain sequential recommendation, CSR）旨在通过挖掘用户在多域混合序列中的行为偏好来为其提供跨域个性化推荐服务. 近年来, 研究人员开始尝试将图卷积网络（graph convolution network, GCN）集成到 CSR 中, 以建模用户和项目之间的复杂关系. 然而, 基于图的 CSR 方法

收稿日期: 2024-03-16; **修回日期**: 2024-04-29
基金项目: 国家自然科学基金项目(62072288); 山东省泰山学者计划项目(tsqn202211154); 山东省自然科学基金项目(ZR2022MF268)
This work was supported by the National Natural Science Foundation of China (62072288), the Taishan Scholar Program of Shandong Province (tsqn202211154), and the Natural Science Foundation of Shandong Province (ZR2022MF268).
通信作者: 赵中英(zzysuin@163.com)

大多通过复杂的结构来捕捉用户在多个域中的序列行为模式，这导致其通常具有较高的计算复杂度和较大的内存开销，限制了模型在资源受限设备上的应用. 此外，已有的轻量级图跨域序列推荐方法认为，应该采用单层聚合协议（single layer aggregating protocol, SLAP）来学习跨域序列图（cross-domain sequential graph, CSG）上的嵌入表示. 基于这种协议的图卷积网络，能够规避多层聚合协议所带来的额外跨域噪声，但却难以捕捉域内的高阶序列依赖关系. 为了解决上述挑战，提出了一种轻量级的三分支图外部注意力网络（tri-branches graph external attention network, TEA-Net）. 具体而言，TEA-Net 首先将原始 CSG 分为域间以及域内序列图，并设计了一种并行的三分支图卷积网络结构来学习图中的节点表示. 该结构能够以较低的计算开销，在不引入额外跨域噪声的条件下，学习域间的低阶协同过滤关系和域内的高阶序列依赖关系. 其次，在三分支结构的基础上，提出了一种改良的外部注意力（external attention, EA）组件，该组件移除了 EA 中的非线性通道，使其能够以更低的开销挖掘项目序列依赖关系并将注意力权重在多个分支上共享. 在 2 个真实数据集上进行了广泛的实验来验证 TEA-Net 的性能表现. 与 10 种最先进的 CSR 方法相比，TEA-Net 在轻量化性能和预测精度方面均取得了更好的结果.

关键词 推荐系统；轻量化跨域序列推荐；外部注意力；图神经网络；协同过滤

中图法分类号 TP391

序列推荐（sequential recommendation, SR）任务旨在从用户的历史交互行为序列中捕捉其动态行为偏好，以建模用户长期的偏好演化过程[1]. 这一研究方向已经引起了大量的关注，相关序列推荐模型也被广泛应用于电子商务、信息检索和移动服务等诸多领域[2-3]. 然而，SR 方法在面临数据稀疏或冷启动推荐场景时，由于缺乏足够的数据来学习用户的序列偏好，其性能往往不尽如人意[4]. 因此，跨域序列推荐（cross-domain sequential recommendation, CSR）方法引发了研究者的关注. 与传统 SR 方法相比，CSR 方法能够借助其他领域（源域）中丰富的辅助信息来解决本域（也称目标域）中的数据稀疏性问题[4]. 在 CSR 场景中，由于用户需要同时在源域与目标域中存在历史交互行为，故研究者们通常将此类用户称作重叠用户. 于是，CSR 可以被总结为：通过建模重叠用户在多个域中的混合行为序列来捕捉其跨域序列偏好的任务.

早期的 CSR 方法倾向于采用基于循环神经网络（recurrent neural network, RNN）的结构[5-6]来捕捉混合序列中的用户偏好. 然而，此类方法对跨域实体之间关系的关注不足，同时难以建模长序列中的用户偏好演化过程. 随后，注意力机制逐渐取代了 RNN，成为了实现 CSR 的热门方法. 该机制利用 Attention 或 Self-Attention 机制来筛选和捕捉混合序列中的序列偏好. 例如，DAT-MDI[7] 和 DASL[8] 采用双重注意力网络来筛选用户的域间序列偏好并共享跨域知识. 这类方法能够充分地考虑项目在混合序列中的全局依赖关系，却难以挖掘重叠用户与跨域项目间的结构化信息. 为了解决这一问题，学者们尝试把 GCN 与 Attention 相结合，将 CSR 任务转移到基于跨域序列图（cross-domain sequential graph, CSG）的非欧式空间中进行解决[9-10]，通过跨域图卷积操作在源域和目标域间传递细粒度的域知识，同时采用自注意力机制来捕捉混合序列中项目之间的全局相关性. 上述基于图的方法已经在 CSR 任务上取得了卓越的性能表现，但仍然面临 2 个挑战：

1）高昂的计算开销. 通过集成自注意力机制来挖掘序列偏好的图推荐方法，大多具有平方级计算复杂度和庞大的参数规模. 这导致此类 CSR 模型难以部署到资源受限的边缘设备，例如，智能手机上[11]. 因此，如何在保证预测精度的前提下提供轻量化的跨域序列推荐成为一项新兴且充满挑战性的任务. 早期的推荐系统轻量化方法[12]，大多关注结构的优化. 例如，wang 等人[13] 提出将用户特征划分为静态和动态特征，并分别采用简化的 GCN 和注意力机制来建模这 2 类特征以达到高效的目的. 然而，文献 [13] 方法并未对图编码器本身进行优化，并且当用户交互数量变得十分庞大且包含跨域信息时，其轻量化性能依旧十分有限. 另一种可行的思路是针对序列编码器展开优化. 例如，LEA-GCN[14] 采用一种双通道外部注意力（external attention, EA）组件来取代基于自注意力的序列编码器. EA 可以通过外部记忆单元来简化注意力计算的时间复杂度，因此在 CSR 任务上实现了较好的轻量化性能. 然而，该组件包含的协同过滤通道仍存在大量的非线性加权设计，这些设计的存在导致该模块仍需要维持较高的计算开销.

2)不恰当的层聚合策略. 众所周知，图神经网络大多采用多层聚合协议(multi layers aggregating protocol, MLAP)来进行节点表示学习. 然而，在CSR场景中，由于源域的高阶序列依赖关系中带有大量的“域-独特”信息，当这些信息通过多层聚合协议经由重叠用户节点传递到目标域的项目节点中时，会产生额外的跨域噪声. 因此，LEA-GCN[14]采用单层聚合协议(single layer aggregating protocol, SLAP)，即仅通过聚合跨域图上的一阶近邻关系来学习节点的嵌入表示. 这样的策略一方面降低了因多层图卷积所带来的高计算复杂度，另一方面也帮助GCN规避了因高阶邻域聚合所带来的跨域噪声干扰. 然而，文献[14]方法无法挖掘单个域内项目间的高阶序列依赖关系. 因此，基于SLAP策略的跨域图序列推荐模型难以建模用户的长期偏好演化过程.

为了解决上述挑战，本文提出一种轻量级的三分支图外部注意力网络(tri-branches graph external attention network, TEA-Net). 具体而言，首先基于混合序列构建包含跨域交互信息的域间序列图(inter-domain sequential graph, ITSG)以及仅包含2个域内项目交互的域内序列图(inner-domain sequential graph, INSG). 然后，提出一个并行的三分支图卷积网络，该网络能够基于不同的层聚合策略来同步地学习ITSG与INSG上的节点表示. 在不引入额外的跨域噪声的条件下，TEA-Net能够以较低的内存开销来同时聚合域间的一阶近邻关系和域内的高阶序列依赖关系. 进一步地，提出一个改进的外部注意力模块，该模块能够以较低的计算复杂度为项目节点进行序列编码，并将注意力权重在不同分支间共享. 此外，为了增强不同分支间节点表示的一致性，TEA-Net还采用对比学习策略对三分支的节点表示进行细化. 最后，在2个大规模真实数据集上的广泛实验证实了CSR方法的有效性.

本文的贡献主要体现在4个方面：

1)提出了一种基于三分支图外部注意力网络的轻量化跨域序列推荐方法.

2)提出了一种改进外部注意力机制. 该机制能够基于较低的计算开销来捕捉项目在序列中的全局相关性并将注意力权重在各分支间共享.

3)设计了一种并行的三分支图卷积网络. 它能够采用灵活的层聚合策略来同时考虑域间的一阶协同过滤关系和域内的高阶序列依赖关系.

4)在2个真实数据集上进行实验，结果表明，与当前最先进的CSR方法相比，TEA-Net具有更高的预测准确性和更低的计算开销.

1 相关工作

本节将从序列推荐、跨域序列推荐和轻量级推荐3个方面来介绍当前的研究现状.

1.1 序列推荐

序列推荐(sequential recommendation, SR)系统旨在通过对用户的历史交互序列进行建模[3]，来捕捉用户行为模式的演变[2]. 已有的SR研究大致可以分为2类，即传统的基于机器学习的方法[15-18]以及基于深度学习的方法[19-21]. 在早期基于机器学习的SR方法中，邢春晓等人[18]提出采用基于协同过滤的序列加权模型来同时考虑基于时间的权重和基于项目相似度的权重. 随后，基于矩阵分解与马尔可夫假设的方法被用来捕获用户的高阶序列行为模式[15]. 文献[15-18]方法在序列推荐的早期探索中发挥了重要的作用. 然而，上述方法大都基于浅层的网络结构，难以捕捉用户深层次的兴趣偏好.

随着深度神经网络的发展，研究人员已经将注意力转移到了基于深度学习的SR方法上来. 例如，HRNN[22]作为一种基于循环神经网络的序列推荐方法，通过一种分层GRU结构在会话序列间进行信息传递，显著提升了SR的预测性能. 然而，此类方法在建模长序列时容易发生梯度消失的问题，并且由于门控机制的存在，算法的结构和时间复杂度都相对较高[23]. 因此，学者们将目光转移到了基于图神经网络(GNN)的方法[24-25]、基于Transformer的方法[26-28]以及基于自我监督的方法[20-21,29]. 例如，GES-SASRec[25]通过融合项目序列关系和基于项目属性的语义关系构建了一种新颖的混合项目图，并通过对混合项目图进行图卷积，以生成平滑的项目节点嵌入. RAISE[28]作为一种基于Transformer的方法，能够基于全局项目关系来分析潜在的用户意图并捕捉候选项目与目标用户之间的特定关系. 此外，zhao等人[21]提出通过对比学习策略增强用户序列表示学习的方法，通过强调序列中顺序关系的一致性来获得对时间间隔不敏感的用户表示. 上述基于深度神经网络的序列推荐方法在SR任务上展现出强大的性能优势，然而这些方法在应对SR的数据稀疏性和推荐系统冷启动问题方面仍存在挑战.

1.2 跨域序列推荐

通过引用和共享来自其他领域的知识，CSR方法可以有效缓解SR方法面临的数据稀疏性和冷启

动问题[30]. 在早期 CSR 的探索过程中，学者们曾采用基于 RNN 的网络结构来实现域间知识的共享，例如 π-Net[5] 和 PSJNet[6]. 这类方法能够利用并行的 RNN 网络结构在不同的域间共享用户偏好信息，并同时通过长短时记忆建模域中的序列偏好. 然而，相比于传统序列推荐场景，CSR 场景由于需要同时考虑重叠用户在多个域中的交互，跨域混合序列中包含的交互规模通常更加庞大，这导致 RNN 的性能更加受限. 为了应对该挑战，基于注意力机制的 CSR[7-8,31] 开始涌现，例如，SCRM[31] 首先通过多层交叉映射感知网络共享域间知识，然后通过自注意力机制捕捉共享知识中的全局序列偏好. DASL[8] 则通过双重自注意力网络来对齐用户在 2 个域中的序列偏好，并为其生成跨域推荐结果. 这类方法能够更精确地捕捉不同域序列中的高阶项目依赖关系，但往往伴随着较高的计算复杂度和时间开销. 随后，GCN 逐渐成为了解决 CSR 问题的热门选择[4,9,32]. 这类方法普遍采用 CSG 的方式来表示跨域序列推荐场景中复杂的用户与项目关系[10]. 例如，DA-GCN[9] 采用图协同过滤编码器来学习 CSG 上的跨域表示，并通过注意力机制来聚合用户的序列偏好. TRaGCF[32] 提出一种基于时间和关系感知的图注意力算法，该方法能够挖掘用户高阶行为模式并利用跨域用户行为模式实现跨域知识的双向迁移. 然而，随着图模型结构复杂性的上升，基于图的 CSR 方法的参数规模和内存开销也渐渐超过了大多数传统深度学习模型，导致其难以被部署到资源受限的边缘设备上.

1.3 轻量级推荐

为了简化推荐系统的结构并提升其在多种设备上应用的便捷性，轻量级推荐系统的概念[12] 引起了广泛关注. 在早期的探索过程中，文献 [33] 提出采用对特征嵌入剪枝的方式优化表示学习，该方法采用一种可学习的阈值来自适应地调节嵌入张量的维度大小. 另一观点则认为可以通过压缩输入的原始协同矩阵的方式来实现轻量化推荐的目标[3,11-12]. 然而，无论是剪枝过程还是矩阵压缩过程，通常都伴随着一定程度的特征损失，甚至丢失关键的项目间的顺序依赖关系. 因此，部分研究开始致力于改良模型的结构复杂度，例如 LightGCN[34]，通过移除传统 GCN 中的特征变换和非线性激活单元，在几乎不影响精度的前提下极大程度地降低了图推荐系统的计算复杂度. LightGCAN[13] 则将原始高维用户特征划分为静态特征和动态交互特征，并采用多个编码器并行的结构来分别进行建模，以达成高效推荐的目标.

尽管在传统推荐任务上已经存在许多轻量化的探索，然而在轻量化序列推荐的研究领域仍然留有大量空白. LSAN[3] 作为提出轻量化序列推荐思想的工作，采用了低秩组合嵌入来优化模型的输入并提出了一种改良的双注意力网络来生成序列推荐. 这种网络结构去除了传统多头自注意力中的冗余，同时保留了捕获长期和短期（即全局和局部）项目依赖性的能力. 随后，LEA-GCN[14] 作为目前仅有的跨域序列推荐工作，在基于 LightGCN 的 CSR 模型基础上提出了采用 SLAP 来简化跨域节点表示学习过程的观点. 此外，文献 [14] 还介绍了一种双通道的外部注意力组件来补充挖掘节点嵌入中的序列偏好. 然而，SLAP 可能会导致 CSR 模型忽略域内的高阶节点序列依赖关系和高阶协同过滤关系. 此外，外部注意力网络中的协同过滤通道仍具有较高的计算复杂度，不利于实现序列表示学习阶段的轻量化目标. 因此，在 CSR 场景下的相关的轻量化推荐方法仍需进一步探索.

2 TEA-Net 框架

本节首先介绍了 TEA-Net 中的符号定义和 CSR 的任务定义，随后对 TEA-Net 模型的每个组件和结构设计进行了详细描述.

2.1 符号与任务定义

CSR 任务的目标是通过对重叠用户在 2 个域中的历史行为序列进行建模，进而分别在每个域上为其生成下一个感兴趣的推荐结果[4]. 为了清晰地描述这一任务，本文首先将重叠用户的集合定义为 $\mathcal{U}=\{u_1,u_2,\cdots,u_k,\cdots,u_p\}(k=1,2,\cdots,p)$. 这些用户均同时在 A 域和 B 域中存在多个历史交互. 其次，A 域的项目集合 $\mathcal{A}=\{A_1,A_2,\cdots,A_i,\cdots,A_m\}(i=1,2,\cdots,m)$，其中 $A_i\in\mathcal{A}$ 表示 A 域中的项目. 而 B 域的项目集合定义为 $\mathcal{B}=\{B_1,B_2,\cdots,B_j,\cdots,B_n\}(j=1,2,\cdots,n)$，其中，$B_j\in\mathcal{B}$ 则表示 B 域中的项目. 此外，由于用户混合序列包含的交互数目通常较多，很难直接作为输入被应用到实际训练中. 因此，本文将重叠用户的原始混合序列进行拆分，令 $\mathcal{S}_A=\{S_1^A,S_2^A,\cdots,S_k^A,\cdots,S_p^A\}$ 和 $\mathcal{S}_B=\{S_1^B,S_2^B,\cdots,S_k^B,\cdots,S_p^B\}$ 分别表示重叠用户在 A 域和 B 域中的交互序列. $S_k^A=\{A_1,A_2,\cdots,A_i,\cdots,A_{t_{ka}}\}$ 表示重叠用户 u_k 在 A 域中的历史行为序列，其中，t_{ka} 为 S_k^A 的序列长度. 类似地，$S_k^B=\{B_1,B_2,\cdots,B_j,\cdots,B_{t_{kb}}\}$ 则表示 u_k 在 B 域中的交互行为序列，其中，t_{kb} 表示 S_k^B 的序列长度. 在拆分

过程中，项目在混合序列中的相对位置也被记录了下来，以便在表示学习过程中补充缺失的混合序列位置关系. 以S_k^A为例，其项目在原始混合序列中的相对位置可以被表示为$P_k^A=\{p_1,p_2,\cdots,p_{t_{ka}}\}$.

因此，在A域中的候选项被推荐给重叠用户u_k的概率可以形式化为：

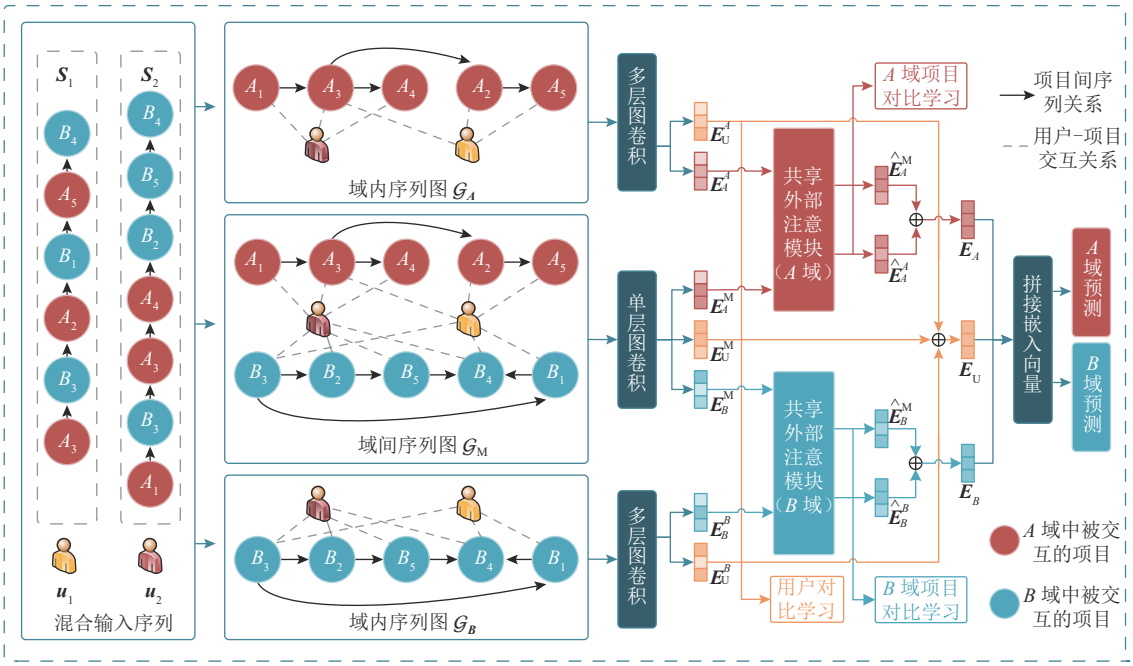
$$P\left(A_{i+1}|S_k^A,P_k^A,S_k^B,P_k^B,u_k\right)\sim f_A\left(S_k^A,P_k^A,S_k^B,P_k^B,u_k\right),\quad(1)$$

$$P\left(B_{j+1}|S_k^B,P_k^B,S_k^A,P_k^A,u_k\right)\sim f_B\left(S_k^B,P_k^B,S_k^A,P_k^A,u_k\right),\quad(2)$$

其中，$P\left(A_{i+1}|S_k^A,P_k^A,S_k^B,P_k^B,u_k\right)$表示在考虑了$u_k$的历史交互序列的前提下，将$A_{i+1}$推荐给用户的概率. $f_A\left(S_k^A,P_k^A,S_k^B,P_k^B,u_k\right)$则表示用于估计该概率分布的似然函数. 同理，在$B$域中也有着相似的定义.

2.2 TEA-Net 模型概述

本节设计了轻量级三分支图外部注意力网络TEA-Net，如图 1 所示，其核心思想是在不影响模型预测精度的前提下，实现一种轻量化的基于图的跨域序列推荐方法. TEA-Net 包含 3 个模块：序列图的构建、节点表示学习以及序列表示学习.



注：u_1与u_2分别表示 2 个不同重叠用户，而S_1和S_2则分别表示他们的混合交互序列.

Fig. 1 The architecture diagram of TEA-Net

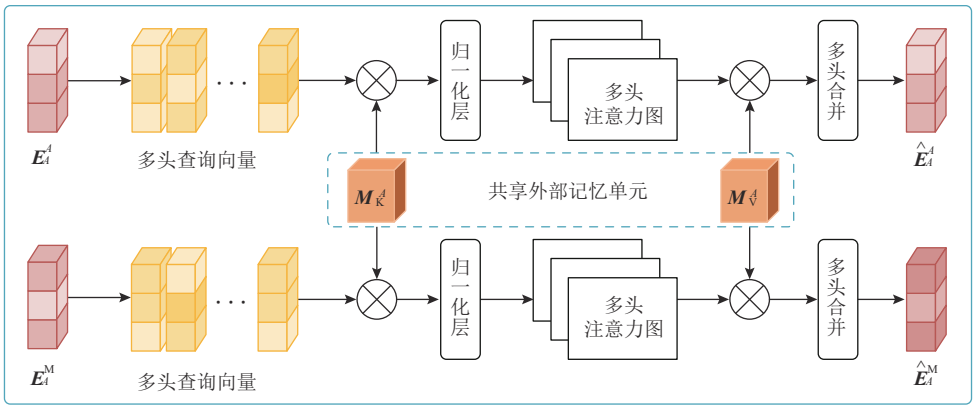
图 1 TEA-Net 的架构图

1)序列图的构建. 传统的基于图的 CSR 方法通常采用构建单个 CSG 的方式来表示重叠用户与跨域项目间的复杂联系. 然而，源域的高阶序列依赖关系中包含着大量的“域-独特”属性信息，当这些信息通过 MLAP 传递到目标域的项目节点中时，会产生跨域噪声影响[14]. 而 SLAP 聚合协议又会导致图中单个域内高阶关系挖掘的不足，需要引入额外的结构来解决. 为此，本文提出了构建多个序列图来建模用户混合行为序列的思想. 如图 1 左侧部分所示，TEA-Net 首先构建了一张完整的域间序列图$\mathcal{G}_M$，图中包含了用户在 2 个域中的全部交互以及 2 个域中全部项目间的序列依赖关系. 其次，构建了 2 张域内序列图$\mathcal{G}_A$与$\mathcal{G}_B$，分别囊括了A域和B域中的用户-项目交互关系以及各自域内的序列依赖关系. 采用这种构图策略的优势在于：首先可以根据 CSR 任务的特点，灵活地在不同的序列图上采取不同的层聚合策略. 其次由于不同的序列图采样自相同的混合行为序列，因此无须增加额外的可学习参数. 最后基于相同混合序列采样得到的不同序列图，有利于学习节点在不同视图间的结构一致性，这为采用对比学习精炼节点嵌入提供了必要条件.

2)节点表示学习. 在构图的基础上设计了一个并行的三分支图卷积网络来学习不同序列图中的节点表示. 由于$\mathcal{G}_A$与$\mathcal{G}_B$中不包含跨域信息，因此在这 2 个序列图上可以采用传统的 MLAP 聚合策略来充分挖掘单个域内序列中的高阶依赖关系；而在$\mathcal{G}_M$上则可以采用 LEA-GCN[14] 提出的 SLAP 策略来聚合跨域序列图上的一阶近邻关系，以避免带来跨域噪声影

响. 上述三分支结构可以灵活地为不同分支配置合适的层聚合策略，并且这种分而治之的并行策略使3个分支上的节点表示学习过程能够同步进行，这也在一定程度上节约了训练的时间成本. 此外，TEA-Net分别在$\mathcal{G}_A$与$\mathcal{G}_{\mathrm{M}}$，$\mathcal{G}_B$与$\mathcal{G}_{\mathrm{M}}$之间对项目嵌入和用户嵌入应用了对比学习策略. 该策略可以通过对齐域间视图与域内视图中节点结构的一致性来提升节点的表示学习效果.

3)序列表示学习. 为了进一步挖掘项目节点表示中隐藏的序列偏好，TEA-Net在三分支结构上提出了一种改进的外部注意力模块设计来加工项目节点表示. 该模块通过引入2个外部注意力单元，能够以接近线性的计算复杂度来度量项目的全局相关性，并将注意力权重在不同分支间共享. 该模块的优越性主要体现在2个方面：首先，三分支结构保护了跨域序列图中的高阶连通关系，TEA-Net因此可以省略外部注意力模块中的非线性通道. 这不仅简化了模型结构，还提升了注意力得分的计算效率. 其次，TEA-Net将外部注意力模块扩展到了一个共享空间中，如图2所示，使其能够同时处理来自不同分支的节点嵌入，并将注意力权重在分支间共享. 这有利于提高模块的利用率，并进一步降低参数规模.



注：$\boldsymbol{M}_{\mathrm{K}}^{A}$和$\boldsymbol{M}_{\mathrm{V}}^{A}$是2个不同的外部记忆单元.

Fig. 2 The sharing external attention components in domain *A*

图2 在*A*域中共享外部注意力组件

最终，TEA-Net通过拼接来自不同分支的用户和项目嵌入，分别产生在2个域上的跨域推荐结果.

2.3 序列图的构建

TEA-Net构建了3张序列图来完整地表示重叠用户的混合交互序列，即$\mathcal{G}_{\mathrm{M}}$，$\mathcal{G}_A$与$\mathcal{G}_B$. 域间序列图$\mathcal{G}_{\mathrm{M}}$中包含了4类关系，即重叠用户分别在*A*域和*B*域中的交互关系以及2个域中各自的序列依赖关系. 为了形式化地描述这2种输入关系，本文采用了矩阵的形式来描述$\mathcal{G}_{\mathrm{M}}$：

$$\mathcal{G}_{\mathrm{M}}=\begin{pmatrix}\mathcal{M}_{A\to A} & \mathcal{M}_{A\to \mathrm{U}} & \mathbf{0}\\ \mathcal{M}_{\mathrm{U}\to A} & \mathbf{0} & \mathcal{M}_{\mathrm{U}\to B}\\ \mathbf{0} & \mathcal{M}_{B\to \mathrm{U}} & \mathcal{M}_{B\to B}\end{pmatrix},\tag{3}$$

其中，$\mathcal{M}_{A\to A}\in\mathbb{R}^{m\times m}$和$\mathcal{M}_{B\to B}\in\mathbb{R}^{n\times n}$分别表示存储着2个域中项目间序列依赖关系的Laplace矩阵，$\mathcal{M}_{A\to \mathrm{U}}\in\mathbb{R}^{m\times p}$和$\mathcal{M}_{\mathrm{U}\to A}\in\mathbb{R}^{p\times m}$表示存储着用户在*A*域中交互关系的矩阵；相应地，$\mathcal{M}_{\mathrm{U}\to B}\in\mathbb{R}^{p\times n}$和$\mathcal{M}_{B\to \mathrm{U}}\in\mathbb{R}^{n\times p}$表示存储着用户在*B*域中交互关系的矩阵.

与此同时，$\mathcal{G}_A$与$\mathcal{G}_B$则仅包含重叠用户在单个域中的交互关系，以及单个域中的项目间序列依赖关系. 于是，它们可以被表示为：

$$\mathcal{G}_A=\begin{pmatrix}\mathcal{M}_{A\to A} & \mathcal{M}_{A\to \mathrm{U}}\\ \mathcal{M}_{\mathrm{U}\to A} & \mathbf{0}\end{pmatrix},\tag{4}$$

$$\mathcal{G}_B=\begin{pmatrix}\mathcal{M}_{B\to B} & \mathcal{M}_{B\to \mathrm{U}}\\ \mathcal{M}_{\mathrm{U}\to B} & \mathbf{0}\end{pmatrix}.\tag{5}$$

由于，$\mathcal{G}_{\mathrm{M}}$，$\mathcal{G}_A$与$\mathcal{G}_B$的输入来自相同的混合交互序列，因此，除$\mathcal{G}_{\mathrm{M}}$中必要的节点参数之外，$\mathcal{G}_A$与$\mathcal{G}_B$在构图过程中无需申请额外的可学习参数，这3张序列图将在节点表示学习阶段共享项目和用户参数，即使用相同的节点表示空间. 也就是说，上述构图方式仅增加了结构复杂度，而并未导致参数规模的增加. 此外，上述构图方法也为学习跨域序列图中的高阶联系提供了条件.

2.4 节点表示学习

为了以较低的内存开销去同步地聚合域间的一阶近邻关系和域内的高阶序列依赖关系，TEA-Net设置了一个三分支图卷积网络来进行节点表示学习. 对于每个分支，TEA-Net统一采用结构最简的LightGCN[33]

作为图编码器，以保证图卷积运算的轻量化性能. 但对于不同的序列图，则采用了不同的层聚合策略. 在$\mathcal{G}_{\mathrm{M}}$分支，TEA-Net 采用 SLAP 协议来抽取跨域结构知识，这样可以有效避免引入额外的跨域噪声. 而对于$\mathcal{G}_A$和$\mathcal{G}_B$，则采用 MLAP 协议来挖掘单个域上的高阶协同过滤与序列依赖关系. 因此，在$\mathcal{G}_{\mathrm{M}}$分支上的节点表示学习过程可以形式化为：

$$\boldsymbol{E}_{\mathrm{U}}^{\mathrm{M}}=\sum_{u_k\in\mathcal{U}}\left(\sum_{A_i\in S_k^A}\frac{1}{\sqrt{|\mathcal{U}|}\sqrt{|S_k^A|}}\boldsymbol{e}_{A_i\to u_k}+\sum_{B_j\in S_k^B}\frac{1}{\sqrt{|\mathcal{U}|}\sqrt{|S_k^B|}}\boldsymbol{e}_{B_i\to u_k}\right),\tag{6}$$

$$\boldsymbol{E}_A^{\mathrm{M}}=\sum_{S_k^A}^{|S_A|}\left(\sum_{A_i\in S_k^A}\sum_{u_k\in\mathcal{U}}\frac{1}{\sqrt{|\mathcal{U}|}\sqrt{|S_k^A|}}\boldsymbol{e}_{u_k\to A_i}+\sum_{A_i\in S_k^A}\frac{1}{|S_k^A|}\boldsymbol{e}_{A_{(i-1)}\to A_i}\right),\tag{7}$$

$$\boldsymbol{E}_B^{\mathrm{M}}=\sum_{S_k^B}^{|S_B|}\left(\sum_{B_j\in S_k^B}\sum_{u_k\in\mathcal{U}}\frac{1}{\sqrt{|\mathcal{U}|}\sqrt{|S_k^B|}}\boldsymbol{e}_{u_k\to B_j}+\sum_{B_j\in S_k^B}\frac{1}{|S_k^B|}\boldsymbol{e}_{B_{(j-1)}\to B_j}\right),\tag{8}$$

其中，$\boldsymbol{E}_{\mathrm{U}}^{\mathrm{M}}$，$\boldsymbol{E}_A^{\mathrm{M}}$，$\boldsymbol{E}_B^{\mathrm{M}}$分别表示 GCN 从$\mathcal{G}_{\mathrm{M}}$分支上学习到的用户节点表示以及 2 个域内的项目表示；$\boldsymbol{e}_{u_k\to A_i}$与$\boldsymbol{e}_{u_k\to B_j}$表示从用户节点分别聚合到 2 个域项目节点上的信息；$\boldsymbol{e}_{A_i\to u_k}$与$\boldsymbol{e}_{B_i\to u_k}$则分别代表从 A 域和 B 域的项目节点聚合到用户节点上的信息；$\boldsymbol{e}_{A_{(i-1)}\to A_i}$与$\boldsymbol{e}_{B_{(j-1)}\to B_j}$分别表示在 A 域和 B 域中项目间序列依赖关系的聚合.

而在$\mathcal{G}_A$和$\mathcal{G}_B$分支上的节点表示学习过程则采用 MLAP，第 l 层上的图卷积过程可以表示为：

$$\boldsymbol{E}^{(l)}=H_{\mathrm{M}}\left(\boldsymbol{E}^{(l-1)},\mathcal{G}\right),\tag{9}$$

其中，$\boldsymbol{E}^{(l)}$表示第 l 层的节点表示，$\boldsymbol{E}^{(l-1)}$表示前一层的节点表示，$H_{\mathrm{M}}(\cdot)$表示 MLAP 层聚合函数.

在$\mathcal{G}_{\mathrm{A}}$分支上的节点表示学习过程可以形式化为：

$$\boldsymbol{E}_{\mathrm{U}}^{A(l)}=\sum_{u_k\in\mathcal{U}}\sum_{A_i\in S_k^A}\frac{1}{\sqrt{|\mathcal{U}|}\sqrt{|S_k^A|}}\boldsymbol{e}_{A_i\to u_k}^{(l-1)},\tag{10}$$

$$\boldsymbol{E}_A^{A(l)}=\sum_{S_k^A}^{|S_A|}\left(\sum_{A_i\in S_k^A}\sum_{u_k\in\mathcal{U}}\frac{1}{\sqrt{|\mathcal{U}|}\sqrt{|S_k^A|}}\boldsymbol{e}_{u_k\to A_i}^{(l-1)}+\sum_{A_i\in S_k^A}\frac{1}{|S_k^A|}\boldsymbol{e}_{A_{(i-1)}\to A_i}^{(l-1)}\right),\tag{11}$$

$$\boldsymbol{E}_{\mathrm{U}}^{A}=\sum_{l=0}^{\eta}\frac{1}{1+\eta}\boldsymbol{E}_{\mathrm{U}}^{A(l)};\ \boldsymbol{E}_A^{A}=\sum_{l=0}^{\eta}\frac{1}{1+\eta}\boldsymbol{E}_A^{A(l)},\tag{12}$$

其中，$\boldsymbol{E}_A^{A(l)}$与$\boldsymbol{E}_{\mathrm{U}}^{A(l)}$分别表示在$\mathcal{G}_A$分支上第 l 层的项目节点表示与用户节点表示；$\boldsymbol{e}_{A_i\to u_k}^{(l-1)}$与$\boldsymbol{e}_{u_k\to A_i}^{(l-1)}$代表第$(l-1)$层上用户-项目交互信息的传递与聚合；$\boldsymbol{e}_{A_{(i-1)}\to A_i}^{(l-1)}$则代表$(l-1)$层上序列依赖关系的传递与聚合；$\boldsymbol{E}_{\mathrm{U}}^{A}$和$\boldsymbol{E}_A^{A}$则分别表示执行 MLAP 后最终该分支输出的用户节点表示与项目节点表示.

采用相似的方式可以得到 B 域中的$\boldsymbol{E}_{\mathrm{U}}^{B}$和$\boldsymbol{E}_B^{B}$，在此处不再赘述.

随后，TEA-Net 采用对比学习策略，将来自不同分支的节点表示进行对齐，以创造更多的自监督信号来辅助跨域节点表示学习. 对于用户节点表示$\boldsymbol{E}_{\mathrm{U}}^{\mathrm{M}}$，$\boldsymbol{E}_{\mathrm{U}}^{A}$，$\boldsymbol{E}_{\mathrm{U}}^{B}$，本文将其设置为 2 批成对的正样本项目表示集合，即$(\boldsymbol{E}_{\mathrm{U}}^{\mathrm{M}},\boldsymbol{E}_{\mathrm{U}}^{A})$与$(\boldsymbol{E}_{\mathrm{U}}^{\mathrm{M}},\boldsymbol{E}_{\mathrm{U}}^{B})$来分别对齐 A 域和 B 域中的用户行为偏好. 以$(\boldsymbol{E}_{\mathrm{U}}^{\mathrm{M}},\boldsymbol{E}_{\mathrm{U}}^{A})$集合为例，定义$(\boldsymbol{x}_i^{\mathrm{M}},\boldsymbol{y}_i^{A})$为其中的一对正例节点的表示，此时$(\boldsymbol{E}_{\mathrm{U}}^{\mathrm{M}},\boldsymbol{E}_{\mathrm{U}}^{A})$集合上的 InfoNCE 损失$\mathcal{L}_{\mathrm{C}}^{\mathrm{UA}}$计算为：

$$\mathcal{L}_{\mathrm{C}}^{UA}=-\sum_{\boldsymbol{x}_i^{\mathrm{M}}\in\boldsymbol{E}_{\mathrm{U}}^{\mathrm{M}}}^{\boldsymbol{y}_i^{A}\in\boldsymbol{E}_{\mathrm{U}}^{A}}\log\left(\frac{\exp\left(\mathrm{sim}\left(\boldsymbol{x}_i^{\mathrm{M}},\boldsymbol{y}_i^{A}\right)\alpha^{-1}\right)}{\sum_{j=1}^{|\mathcal{U}|}\exp\left(\mathrm{sim}\left(\boldsymbol{x}_i^{\mathrm{M}},\boldsymbol{y}_j^{A}\right)\alpha^{-1}\right)}\right),\tag{13}$$

其中，$\mathrm{sim}(\cdot)$表示相似性度量函数，这里采用内积来计算样本间相似度；α为温度超参数，它影响了对数相似性的软化程度. 在 3.5 节对α的取值进行了探索.

同理，可以采用相同的计算方式从$(\boldsymbol{E}_{\mathrm{U}}^{\mathrm{M}},\boldsymbol{E}_{\mathrm{U}}^{B})$集合中获得$\mathcal{L}_{\mathrm{C}}^{UB}$. 最后，通过合并 3 个分支上的用户节点表示$\boldsymbol{E}_{\mathrm{U}}^{\mathrm{M}}$，$\boldsymbol{E}_{\mathrm{U}}^{A}$，$\boldsymbol{E}_{\mathrm{U}}^{B}$得到最终的跨域用户表示$\boldsymbol{E}_{\mathrm{U}}$：

$$\boldsymbol{E}_{\mathrm{U}}=\boldsymbol{E}_{\mathrm{U}}^{\mathrm{M}}+\boldsymbol{E}_{\mathrm{U}}^{A}+\boldsymbol{E}_{\mathrm{U}}^{B}.\tag{14}$$

而对于项目节点表示，TEA-Net 则首先在表示的基础上采用一种共享外部注意力组件进行了序列级表示学习. 随后，在加工后的项目表示上又采取了与式(13)相同的对比学习策略(具体细节于 2.5 节介绍).

2.5 序列表示学习

基于传统自注意力机制的序列编码器存在 2 个显著的缺陷：1)该机制采用自值(self-value)的组合来细化输入的序列表示，它只考虑序列中项目之间的关系，而忽略了不同序列中项目间的隐含关系[35]. 2)该机制具有高达$O(n^2)$的计算复杂度，导致计算的时间成本较高.

为了以较低的复杂度准确地测量不同项目对序列的贡献，本节引入了 EA 机制[10,36]. 在此之前，该机

制已在图像分类、目标检测和语义分割方面取得了巨大成功[35]. 与采用 EA 的工作不同，如 LEA-GCN[14]，TEA-Net 所提出的序列编码器结构只保留了 EA 最核心的外部记忆单元和多头机制，舍弃了计算复杂度较高的非线性通道设计. 此外，还根据三分支卷积输出的特点设计了一种共享外部记忆单元的结构，使得多个分支上的项目嵌入可以同步地使用外部记忆单元中维护的全局注意力权重.

如图 2 所示，该外部注意力组件被部署在 $\mathcal{G}_A$ 与 $\mathcal{G}_{\mathrm{M}}$ 分支上来同步地捕捉 $\boldsymbol{E}_A^A$ 与 $\boldsymbol{E}_A^{\mathrm{M}}$ 中的项目全局相关性. 与自注意力机制不同，EA 通过 2 个低秩的外部记忆单元 $\boldsymbol{M}_{\mathrm{K}}^A \in \mathbb{R}^{d\times d}$ 与 $\boldsymbol{M}_{\mathrm{V}}^A \in \mathbb{R}^{d\times d}$ 来计算 A 域中项目的全局共享权重，它们在计算过程中担任键矩阵和值矩阵的角色.

为了更好地形式化注意力的计算过程，此处设 $\boldsymbol{E}_{S_k}^A \in \mathbb{R}^{t_{ka}\times d}$ 为 $\boldsymbol{E}_A^A$ 中序列 S_k^A 的表示，而 $P_{S_k}^A \in \mathbb{R}^{t_{ka}\times d}$ 则代表序列中每个项目在原始混合序列中的相对位置编码. 在计算过程中，该组件首先将 $(E_{S_k}^A \oplus P_{S_k}^A)$ 映射为多头查询向量 $Q_{S_k}^A = \left(\boldsymbol{q}_{S_k}^{A\,(1)}, \boldsymbol{q}_{S_k}^{A\,(2)}, \cdots, \boldsymbol{q}_{S_k}^{A\,(t)}, \cdots, \boldsymbol{q}_{S_k}^{A\,(\beta)}\right)$，其中，$\beta$ 为控制注意力头个数的超参数，$\boldsymbol{q}_{S_k}^{A\,(t)}$ 为第 t 个注意力头.

随后，外部注意力运算过程可以形式化为：

$$\boldsymbol{A}^{(t)} = a_{i,j}^{S_k} = Norm\left(\frac{\boldsymbol{q}_{S_k}^{A\,(t)} \boldsymbol{M}_{\mathrm{K}}^{A\,\mathrm{T}}}{\sqrt{d}}\right), \tag{15}$$

$$\hat{\boldsymbol{E}}_{S_k}^{A(t)} = \boldsymbol{A}\boldsymbol{M}_{\mathrm{V}}^A, \tag{16}$$

其中，$\boldsymbol{A}^{(t)} \in \mathbb{R}^{t_{ka}\times d}$ 为注意力矩阵；$a_{i,j}^{S_k}$ 代表着序列 S_k^A 中第 i 个项目与 $\boldsymbol{M}_{\mathrm{K}}^{\mathrm{A}}$ 中第 j 行之间的相关度；$\hat{\boldsymbol{E}}_{S_k}^{A(t)}$ 则表示加权后的序列表示.

于是，最终输出的序列级表示 $\boldsymbol{H}_S$ 可以表示为：

$$\hat{\boldsymbol{E}}_A^A = \sum_{S_k^A}^{t_{ka}} Concat\left(\boldsymbol{H}_{S_k}^{A\,(1)}, \boldsymbol{H}_{S_k}^{A\,(2)}, \cdots, \boldsymbol{H}_{S_k}^{A\,(\beta)}\right)\boldsymbol{W}_1, \tag{17}$$

其中，$\hat{\boldsymbol{E}}_A^A$ 表示加权后的 $\mathcal{G}_A$ 分支上的项目表示.

同理，在 $\mathcal{G}_{\mathrm{M}}$ 分支上也可以根据相同的计算方式获得 $\hat{\boldsymbol{E}}_A^{\mathrm{M}}$.

总体而言，TEA-Net 所采用的共享外部注意力结构存在 4 点优势：1）外部注意力由于采用了 2 个低秩的外部存储单元来计算全局相关性，其计算复杂度 $O(t_{ka}\times d^2)$ 相较于自注意力 $O(t_{ka}^2\times d)$ 有着显著的降低；2）不同分支间共享同一对外部记忆单元有利于对齐来自不同分支上的项目表示；3）避免了为不同分支单独配置外部记忆单元，节约了内存开销；4）继续在不同分支的工作流上保持并行，有助于提升表示学习效率，节约时间成本.

同理，通过在 $\mathcal{G}_{\mathrm{M}}$ 和 $\mathcal{G}_B$ 之间应用共享外部注意力组件，可以得到另一组加权后的项目表示，即 $\hat{\boldsymbol{E}}_B^B$ 与 $\hat{\boldsymbol{E}}_B^{\mathrm{M}}$. 在此基础上，为了增强 A 域和 B 域上项目节点的一致性，TEA-Net 采用了相同的对比学习策略. 在 $(\hat{\boldsymbol{E}}_A^A, \hat{\boldsymbol{E}}_A^{\mathrm{M}})$ 与 $(\hat{\boldsymbol{E}}_B^B, \hat{\boldsymbol{E}}_B^{\mathrm{M}})$ 之间分别采用式（13）所介绍的 InfoNCE 损失，以获得 $\mathcal{L}_C^A$ 与 $\mathcal{L}_C^B$.

最后，TEA-Net 通过合并用户表示以及不同分支上的项目表示来得到最终的 A 域项目表示 $\boldsymbol{H}_A$ 与 B 域项目表示 $\boldsymbol{H}_B$：

$$\boldsymbol{H}_A = Concat\left(\left(\boldsymbol{E}_A^{\mathrm{M}} + \boldsymbol{E}_A^{\mathrm{A}}\right), \boldsymbol{E}_{\mathrm{U}}\right), \tag{18}$$

$$\boldsymbol{H}_B = Concat\left(\left(\boldsymbol{E}_B^{\mathrm{M}} + \boldsymbol{E}_B^{\mathrm{B}}\right), \boldsymbol{E}_{\mathrm{U}}\right). \tag{19}$$

2.6 预测层

在上述表示学习过程结束之后，TEA-Net 分别获得了 2 个域中带有用户信息的序列表示 $\boldsymbol{H}_A$ 和 $\boldsymbol{H}_B$. 为了充分利用跨域知识，本文在预测前对它们进行了特征拼接，并将拼接后的嵌入一起馈送给了预测层：

$$P(A_{i+1}|S_A, S_B) = softmax\left(\boldsymbol{W}_A \cdot [\boldsymbol{H}_A, \boldsymbol{H}_B]^{\mathrm{T}} + \boldsymbol{b}_A\right), \tag{20}$$

$$P\left(B_{j+1}|S_B, S_A\right) = softmax\left(\boldsymbol{W}_B \cdot [\boldsymbol{H}_B, \boldsymbol{H}_A]^{\mathrm{T}} + \boldsymbol{b}_B\right), \tag{21}$$

其中，S_A 与 S_B 分别表示 A 域和 B 域中所有的输入序列，$\boldsymbol{W}_A$ 和 $\boldsymbol{W}_B$ 分别是 A 域和 B 域中所有项目的权重矩阵；$\boldsymbol{b}_A$ 和 $\boldsymbol{b}_B$ 是 2 个域的偏置项. 在先前的 CSR 工作中，π-Net[5] 和 DA-GCN[9] 均采用了联合损失函数来一起优化 2 个域中的参数，但是这样可能会导致 2 个域的损失发生此消彼长的跷跷板现象（see-saw phenomenon）. 因此，TEA-Net 选择分别在 2 个域上计算交叉熵损失，并独立地对它们进行优化：

$$\mathcal{L}_{\mathrm{CE}}^A = -\frac{1}{|\mathcal{S}|}\sum_{S_A, S_B\in\mathcal{S}}\sum_{A_i\in S_A} \log P(A_{i+1}|S_A, S_B), \tag{22}$$

$$\mathcal{L}_{\mathrm{CE}}^B = -\frac{1}{|\mathcal{S}|}\sum_{S_B, S_A\in\mathcal{S}}\sum_{B_j\in S_B} \log P\left(B_{j+1}\middle| S_B, S_A\right), \tag{23}$$

其中，$\mathcal{S}$ 表示 2 个域全部的训练序列.

因此，2 个域的完整损失可以表示为交叉熵损失与对比损失的加和：

$$\mathcal{L}_A = \mathcal{L}_{\mathrm{CE}}^A + \mathcal{L}_{\mathrm{C}}^A + \mathcal{L}_{\mathrm{C}}^{\mathrm{UA}}, \tag{24}$$

$$\mathcal{L}_B = \mathcal{L}_{\mathrm{CE}}^B + \mathcal{L}_{\mathrm{C}}^B + \mathcal{L}_{\mathrm{C}}^{\mathrm{UB}}. \tag{25}$$

与通过有损的扰动或其他破坏图结构的方法所产生自监督信号不同，TEA-Net 通过对齐不同的分支上的嵌入（即强调域间序列图和域内序列图上节点的结构一致性）来进行对比学习. 由于不包含额外的数据增广噪声，这里并未设置其他超参数来测试对比损失的影响，而是采用直接相加的策略让其尽可能地辅助原始的跨域监督学习任务.

3 实验与分析

本文在2个来自真实世界的数据集上进行实验以验证TEA-Net的有效性，重点回答4个研究问题：

1)作为一种轻量化CSR模型，TEA-Net的轻量化性能表现如何？各组件对轻量化效果的贡献如何？

2)与其他先进的基线方法相比，TEA-Net的预测性能表现如何？其轻量化策略是否导致了推荐性能的下降？

3)采用三分支图卷积结构是否有助于节点表示学习？共享外部注意力组件的表现如何？对比学习的效果如何？

4)对比学习损失中的温度系数α对TEA-Net的影响如何？共享外部注意力组件中的多头数目β对性能有多少影响？

3.1 实验设置

1)数据集. 本文在2个真实世界的数据集豆瓣(DOUBAN)[37]和亚马逊(AMAZON)[38]上进行实验，数据集的统计信息如表1所示.

Table 1 Statistics of the Datasets

表1 数据集的统计信息

统计项目	DOUBAN		AMAZON	
	A 域	*B* 域	*A* 域	*B* 域
项目种类	14 636	2 940	126 526	61 362
交互个数	607 523	360 798	1 678 006	978 226
用户个数	6 582		9 204	
训练序列个数	42 062		90 574	
测试序列个数	10 431		14 463	

①DOUBAN. 该数据集收集自中国知名社交媒体平台豆瓣网[37]，其包含了用户在2个域上的历史交互记录，如*A*域(豆瓣电影)和*B*域(豆瓣图书)，这些交互被按照发生的时间顺序以混合序列的形式记录了下来.

②AMAZON.该数据集则是一个经典的产品评论数据集[38]，其包含了亚马逊网站上2个不同平台中用户的评论行为，即亚马逊图书(*A*域)和亚马逊电影(*B*域).

2个数据集中数据的分类和交互个数等统计信息均在表2中给出. 此外，对于2个数据集中所有的混合序列，随机选择了其中80%作为训练序列，其余20%作为测试序列. 并且，本文遵循大多数CSR模型的实验设置[5,9,14]，过滤掉了2个数据集中历史交互次数少于10次的用户以及只被交互了不超过5次的项目.

Table 2 Experiment Results of Comparative Methods on DOUBAN and AMAZON Datasets

表2 对比方法在数据集DOUBAN和AMAZON上的实验结果 %

方法	DOUBAN数据集						AMAZON数据集					
	A 域			*B* 域			*A* 域			*B* 域		
	RC	*MRR*	*NDCG*	*RC*	*MRR*	*NDCG*	*RC*	*MRR*	*NDCG*	*RC*	*MRR*	*NDCG*
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
SASRec	80.56	77.45	46.23	65.77	54.25	38.22	22.75	18.03	11.25	25.46	24.08	16.03
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
SCRM	83.55	81.07	51.83	71.44	59.31	41.30	24.03	18.44	12.56	29.65	27.43	17.96
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
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
TEA-Net	**83.91**	**81.35**	**52.24**†	**76.38**†	**67.58**†	**45.55**†	**25.87**†	**22.27**†	**15.31**†	**34.29**†	**31.18**†	**20.96**†

注：粗体值表示最优结果；“†”标记了与次优结果相比时具有显著提升的结果（经过配对样本t检验，p<0.05）.

2)评价指标. 本文采用3个常用的评价指标*RC*@10, *MRR*@10和*NDCG*@10) [9,39]. 其定义为：

1)*RC*(recall). *RC*是一个衡量系统检索性能的指标，通常用于衡量推荐系统能够召回到的相关项目

的比例. 它的计算方式是相关项目的数量除以所有项目的总数. *RC* 关注的是遗漏召回相关项目的风险，希望尽可能多地找到所有相关的项目.

2) *MRR*(mean reciprocal rank). *MRR* 是衡量推荐系统或检索系统性能的一种常用评价指标，它关注的是排在检索结果列表中第 1 个相关项目的排名(即平均倒数排名). 如果一个用户查询的第 1 个相关文档排在较高的位置，那么 *MRR* 值就会较高. *MRR@N* 的计算公式为：

$$MRR@N = \frac{1}{|Q|}\sum_{q\in Q}\frac{1}{\sum\limits_{k\in[1,N]} R_q(k)},$$

其中，Q表示查询数量，$q\in Q$则表示一个具体的查询，$R_q(k)$ 表示应用查询q时在召回结果集合中的第 k 个项目的排名.

3) *NDCG*(normalized discounted cumulative gain). *NDCG* 是另一种常见的推荐系统评价指标，它同时考虑了项目在召回序列中的排名和项目本身的相关性. 具体而言，它对召回的项目按其排名来计算 *DCG*(discounted cumulative gain)值，然后除以理想情况下(即完美排序的情况下)的 *DCG* 值，以达到归一化效果. *NDCG* 倾向于为排名靠前的相关项目赋予更高的权重. 公式为：

$$NDCG@N = \frac{DCG@N}{IDCG@N},$$

其中，*DCG* 是折现累积获得值，I*DCG* 是理想的折现累积获得值. *DCG* 的计算考虑了排名的折现因子，通常是对召回结果的排名进行指数折现.

3) 基线方法. 为了验证 TEA-Net 的预测性能，本文将其与 3 类共 10 种基线方法进行了对比实验：

①传统推荐方法. 此处采用了 2017 年的 NCF[1]、2019 年的 NGCF[40] 和 2020 年的 LightGCN[34] 三个里程碑式的传统推荐方法作为参考基线方法.

②序列推荐方法. 此处采用了具有代表性的 2016 年的 GRU4REC[2]、2017 年的 HRNN[22] 和 2018 年的 SASRec[41] 三种序列推荐方法作为参考基线方法.

③跨域序列推荐方法. 此处采用了近年来有竞争力的 4 种 CSR 方法作为对比方法，包括 2021 年的 SCRM[31]、2023 年的 PSJNet[6]、2024 年的 TiDA-GCN[10] 和 2023 年的 LEA-GCN[14].

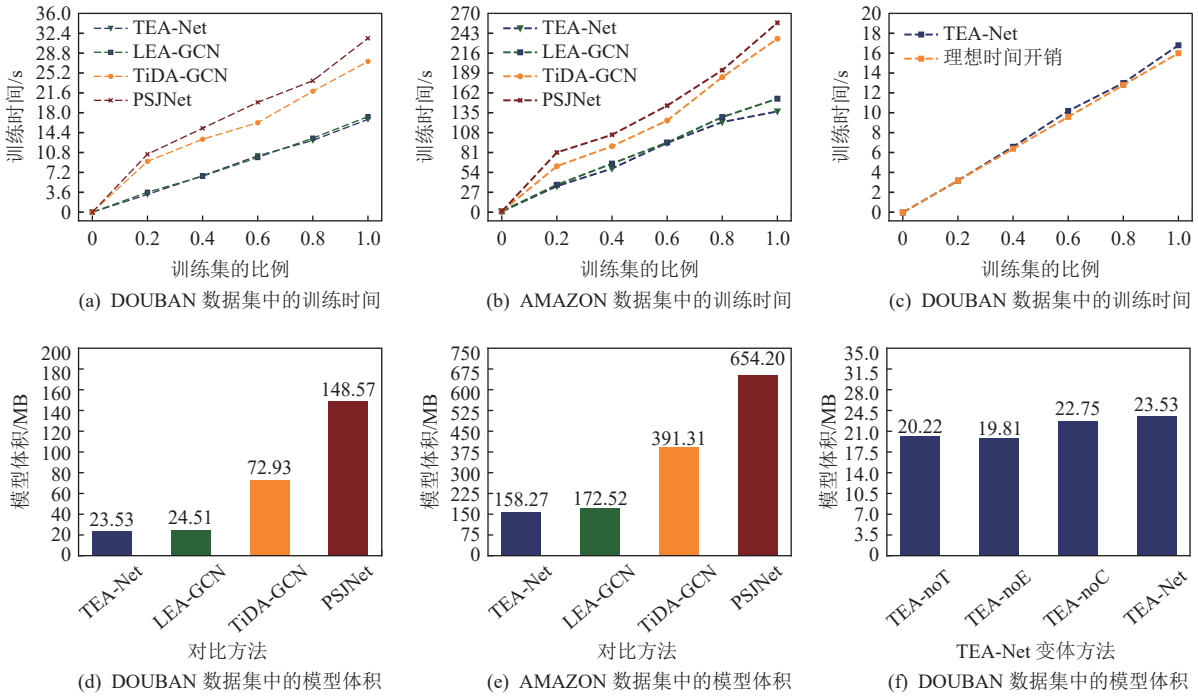
4) 实验设置. TEA-Net 通过 TensorFlow 实现，并通过 NVIDIA RTX 3090 GPU 加速模型的训练过程. 对于实验参数，采用了 Xavier[42] 和 Adam[43] 来分别进行初始化和参数优化. 给定数据集的混合输入序列，TEA-Net 将 *A* 域和 *B* 域最后被观测到的 2 个项目分别视作 2 个域的预测目标(即预测真值). 对于模型训练，TEA-Net 的输入批次大小设置为 256，将丢弃率设置为 0.1，将 *A* 域的学习率设置为 0.005，将 *B* 域的学习率分别设置为 0.004. 对于 TEA-Net 的具体参数，嵌入大小维度被设置为 16，正则化比率被设置为 1E−7. 此外，TEA-Net 在 [0.1, 1] 范围中以 0.1 为步长探索超参数α的影响，以了解温度参数对对比学习性能的影响. 同时，也在$\beta\in\{1, 2, 4, 8, 16\}$范围中探索了对外部注意力组件的影响. 本文在 3.4 节中详细介绍了超参数α和β对试验结果的影响与分析. 此外，为了保证对比结果的公平性，所有基线方法的嵌入大小被统一设置为 16. 值得说明的是，为了使传统推荐基线方法具备可比性，本文在实验过程中将它们的输入统一改为了序列化输入，并分别报告了它们在各个数据集中每个域的实验结果. 而对于序列推荐基线方法，本文则在实验结果中分别报告了它们在每个域中的性能表现. 对于基线方法的其他超参数，统一采用了相应论文中的最佳设置. 为了促进轻量化跨域序列推荐的研究与发展，本文已将 TEA-Net 的相关代码及数据集提交到 GitHub 以实现开源①.

3.2 轻量化性能（针对问题 1）

为了测量模型训练的时间开销，本文分别在数据集 DOUBAN 和 AMAZON 上对 TEA-Net 以及 3 个有代表性的 CSR 基线方法开展了一系列实验. 具体而言，此处按照 0.2, 0.4, 0.6, 0.8, 1.0 的不同训练集比例，通过多次改变输入数据的规模来观察模型所需的训练时间，这有助于测试模型在适应不同规模的训练数据时的表现. 同时，还额外将 TEA-net 的训练时间成本与理想的线性训练时间成本进行了比较，以观察 TEA-Net 的轻量化表现.

其次，本文通过比较 TEA-Net 与其他 3 个具有竞争力的 CSR 基线方法的模型体积(参数规模)来进一步考察了它们的轻量化性能. 最后，还对 TEA-Net 中包含的各组件的模型体积进行了探索，这是为了展示各组件对模型轻量化性能的贡献. 上述实验结果被汇总在图 3 中，对于图 3 中的观测结果存在 5 点分析：1)从图 3(a)(b)中可以发现，TEA-Net 的训练时间开销远远低于 TiDA-GCN 和 PSJNet，这表明 TEA-Net 相比于其他常规 CSR 模型具有更好的训练效率，并且具备扩展到大规模数据集的潜力. 2)TEA-Net 与

① https://github.com/ZZY-GraphMiningLab/TEA-Net



(a) DOUBAN 数据集中的训练时间 (b) AMAZON 数据集中的训练时间 (c) DOUBAN 数据集中的训练时间

(d) DOUBAN 数据集中的模型体积 (e) AMAZON 数据集中的模型体积 (f) DOUBAN 数据集中的模型体积

Fig. 3 Time consumption and the model volume of TEA-Net compared with LEA-GCN, TiDA-GCN and PSJNet

图 3 TEA-Net 与 LEA-GCN，TiDA-GCN 和 PSJNet 相比的时间消耗和模型体积

轻量化跨域序列推荐模型 LEA-GCN 有着较为接近的训练时间开销. 这证明 TEA-Net 能够在保持相似的轻量化性能的同时实现更高的预测精度. 3)从图 3(c)中可以观察到 TEA-Net 的表现非常接近线性增长，这体现了 TEA-Net 在面对大规模数据时的稳定性. 4)通过观察图 3(d)(e)可以发现，TEA-Net 与 3 种基线方法相比拥有最小的模型体积. 其中，TEA-Net 的模型体积小于 LEA-GCN，这进一步显示了 TEA-Net 在轻量化目标上的优越性. 5)与 TEA-Net 相比，移除了对比学习组件的变体模型 TEA-noC、改用单分支图卷积模块的变体模型 TEA-noT 以及移除了共享外部注意力模块的变体模型 TEA-noE 在模型体积方面并没有显著的差异，这恰恰说明每个模块所需的模型体积均相对较小，体现了各组件的轻量化优势.

3.3 预测准确性（针对问题 2）

为了验证 TEA-Net 的预测性能，本文将其与 3 种类别标签下的 10 个基线方法进行了综合比较. 表 2 显示了 TEA-Net 以及其他 3 类基线方法在 2 个数据集上的预测性能表现. 对实验结果的观察与分析得出：1)TEA-Net 在 2 个数据集上的全部评价指标上均取得了最优的性能表现，并且与次优结果相比在大多数评估指标上都具有显著的提升. 这表明 TEA-Net 中采用的轻量级策略没有对预测性能产生消极影响，甚至一定程度上提高了 CSR 模型的预测性能. 2)跨域序列推荐方法普遍优于其他类别中的基线方法(即传统推荐方法和序列推荐方法)，这证明了同时建模用户的序列偏好和考虑跨领域特征的重要性. 3)TEA-Net 优于所有的 CSR 基线方法，表明所提出的三分支图外部注意力网络在建模用户混合序列的问题上具有更好的预测性能. 4)TEA-Net 的性能更优于 LEA-GCN，这证明了针对节点表示学习方法以及序列表示学习方法的改进不仅在轻量化性能上表现更好，并且在预测性能上也要更优于现有的轻量化跨域序列推荐方法.

3.4 消融实验（针对问题 3）

本节以 DOUBAN 数据集为例开展了消融实验，以探索 TEA-Net 中不同组件的贡献. 由于 TEA-Net 的组件间耦合度较高，因此需要通过与 6 组模型变体对比来验证各组件的有效性. 变体模型的定义为：1)TEA-S_{LAP} 表示在其他组件不变的情况下，对三分支图卷积统一采用 SLAP 的变体；2)TEA-noC 移除了 TEA-Net 中的对比学习组件(包括项目对比学习和用户对比学习)；3)TEA-noT 表示采用单分支图卷积的变体模型，请注意，由于 TEA-Net 的对比学习策略是在不同分支的节点嵌入间展开的，因此，在该变体中对比学习的作用也被移除了；4)在 TEA-noS 变体中，共享外部注意力组件中的共享机制被取消，即分别给每个分支单独设置一组外部记忆单元来计算注意

力得分；5)TEA-noE 则表示彻底移除共享外部注意力组件的变体；6)TEA-noA 表示移除了所有组件的 TEA-Net 变体，此时模型退化为采用跨域序列图作为输入的 LightGCN 网络.值得说明的是，相较于仅使用单个域信息的传统 LightGCN，该变体因保留了跨域嵌入拼接的设计而具备更好的性能.

本文基于对表 3 的观察结果进行了 6 点分析：1)TEA-Net 的性能表现优于 TEA-S_{LAP}，这表明在三分支结构下采取不同的层聚合策略的必要性. 同时，也证明了挖掘域内高阶序列依赖关系与高阶协同过滤关系的重要性. 2)TEA-Net 的性能优于 TEA-noC，这表明在不同的分支间应用对比学习策略的确有助于对齐跨域节点表示. 3)TEA-Net 与 TEA-S_{LAP} 的表现优于 TEA-noT，进一步证明了采用三分支图卷积网络结构有助于提高模型在 CSR 任务上的表现. 4)TEA-Net 的表现优于 TEA-noS，这证明了外部记忆单元在共享外部注意力组件中的重要性，体现了该设计在辅助对齐不同分支之间项目嵌入的作用. 5)TEA-Net 以及其他全部变体（除 TEA-noA 之外）的表现均显著优于 TEA-noE，这证明了共享外部注意力组件对节点序列级表示学习过程的贡献，同时也证明了该组件在建模用户序列行为模式过程中的贡献. 6)TEA-Net 以及其他 5 个变体的性能均显著优于 TEA-noA，这证明 TEA-Net 的所有组件均是有效的.

Table 3 Results of Ablation Studies of TEA-Net on DOUBAN Dataset

表 3 TEA-Net 在 DOUBAN 数据集上的消融实验结果 %

本文方法及其变体	*A* 域			*B* 域		
	RC	*MRR*	*NDCG*	*RC*	*MRR*	*NDCG*
TEA-S_{LAP}	83.41	81.19	52.08	76.14	66.28	45.12
TEA-noC	83.26	80.17	51.58	75.44	65.82	44.03
TEA-noT	83.12	80.03	51.16	74.86	64.17	43.58
TEA-noS	83.16	80.24	51.21	75.05	64.60	43.74
TEA-noE	82.66	79.43	50.25	73.48	63.09	41.95
TEA-noA	82.12	78.10	49.64	71.28	59.43	41.28
TEA-Net	**83.91**	**81.35**	**52.24**	**76.38**	**67.58**	**45.55**

注：粗体值表示最优结果.

3.5 参数分析（针对问题 4）

本节介绍了 TEA-Net 中的 2 个关键超参数α与β. α为对比损失计算过程中使用的温度系数，该系数控制着负样本对正样本的相对影响力，决定了在计算对比损失时样本间相似度所赋的权重. 而β则控制了共享外部注意力组件中多头机制所使用的注意力头个数. 为了探究上述超参数对 TEA-Net 性能的影响，本文以 DOUBAN 数据集为例，在 [0.1, 1] 的区间内以 0.1 为步长进行了 10 组超参数实验来验证α对模型性能的影响；同样地，在$\beta\in\{1, 2, 4, 8, 16\}$的参数设置下也探索了$\beta$对模型性能的影响. 图 4 显示这 2 种超参数在不同的设置下 TEA-Net 的性能表现. 基于观察结果的分析有：

1)当α=0.4 时，TEA-Net 的性能达到最优，而当α趋向 0.1 或 1 时，模型的性能均处于波动下降的状态. 当α趋向于 0.1 时模型会更加强调细粒度信息的捕捉，但同时也可能使得模型对于噪声或者干扰更加敏感. 而当α趋于 1 时，会降低模型对样本间相似度差异的感知，同样影响了模型的性能. 因此，只有采用适中的α，才能最大限度的发挥 TEA-Net 中对比学习策略的优势.

2)当注意力头个数β=4 时，TEA-Net 达到了最优性能. 当继续增加β时，性能反而出现了下降. 这一结果说明，较少的β可能更有利于模型在 CSR 场景中的表达，由于一段时间内用户的兴趣通常不会存在太大差异，因此，仅需要少量的β即可模拟用户有限的兴趣个数. 这一结果符合现实世界的客观规律.

4 结 论

本文提出了一种轻量级的基于 GCN 的 CSR 方法 TEA-Net. 该方法首先针对跨域序列推荐场景的特点，设计了一种并行的三分支图卷积结构. 该结构优化了过去基于单分支的多层图卷积网络，并通过应用不同的层聚合方法以及并行策略来降低内存开销. 其次，TEA-Net 在三分支网络基础上提出了一种改良的共享外部注意力组件. 该组件能够以更低的时间复杂度来进一步挖掘项目节点表示中的序列依赖关系. 此外，在未引入额外输入和扰动的情况下，采用了对比学习策略来增强不同分支间的节点表示学习能力. 本文在 2 个真实世界数据集上进行实验，分别对模型的轻量化性能、预测性能、各组件贡献以及超参数进行了评估，实验结果证实了所提出的轻量化 CSR 方法的优越性.

作者贡献声明：张劲羽负责模型总体设计和论文撰写；马晨曦负责算法实现与论文修改；李超指导论文的修改；赵中英指导模型设计、论文的撰写和修改.

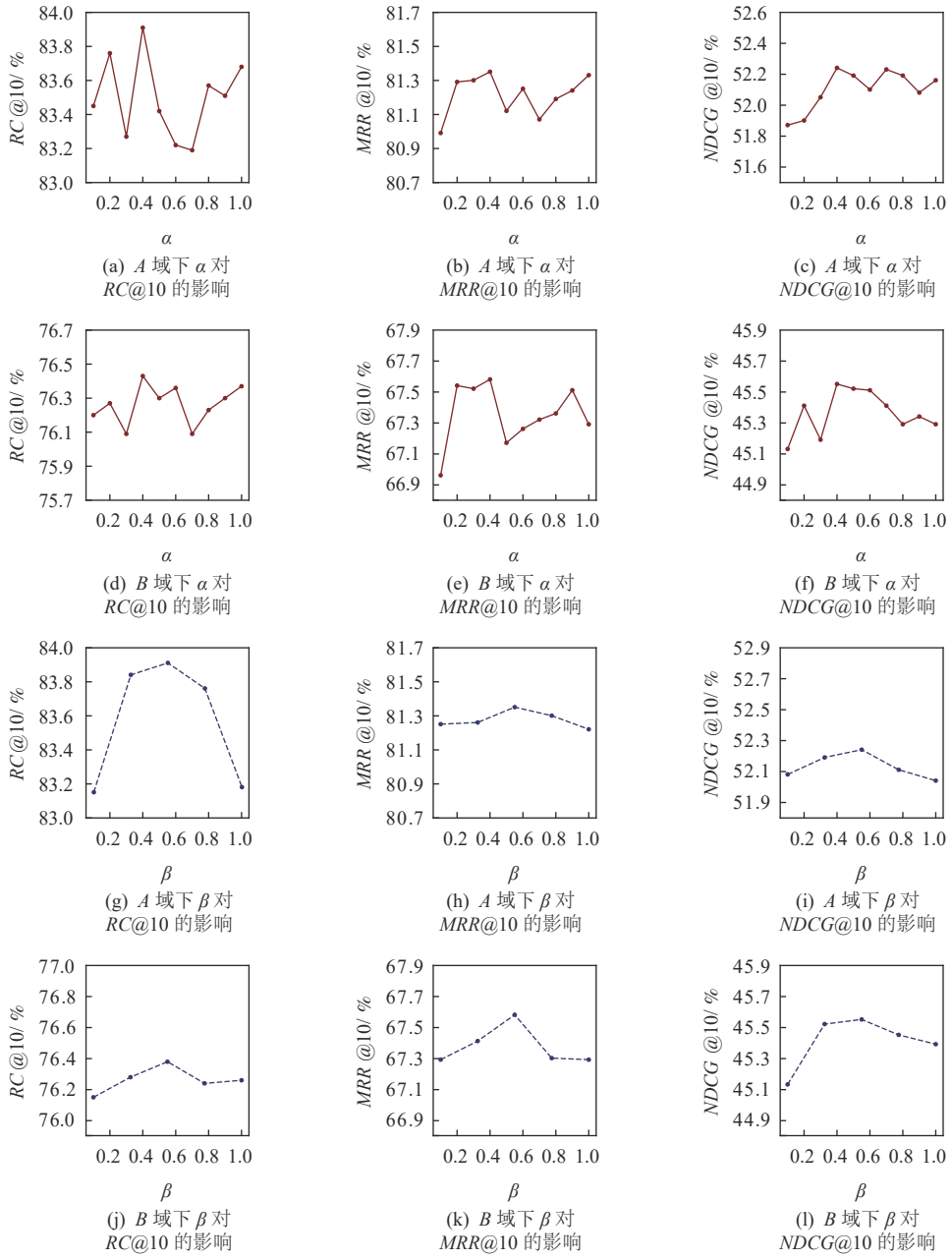


Fig. 4 Impact of hyper-parameters α and β on DOUBAN dataset

图 4 超参数α和β在 DOUBAN 数据集上的影响

参 考 文 献

[1] He Xiangnan, Liao Lizi, Zhang Hanwang, et al. Neural collaborative filtering[C]//Proc of the 26th Int Conf on World Wide Web. New York: ACM, 2017: 173−182

[2] Hidasi B, Karatzoglou A, Baltrunas L, et al. Session-based recommendations with recurrent neural networks[C]//Proc of the 4th Int Conf on Learning Representations, New York: penReview. net, 2016: 1−10

[3] Li Yang, Chen Tong, Zhang Pengfei, et al. Lightweight self-attentive sequential recommendation[C]//Proc of the 30th ACM Int Conf on Information and Knowledge Management. New York: ACM, 2021: 967−977

[4] Zheng Xiaolin, Su Jiajie, Liu Weiming, et al. DDGHM: Dual dynamic graph with hybrid metric training for cross-domain sequential recommendation[C]//Proc of the 30th ACM Int Conf on Multimedia. New York: ACM, 2022: 471−481

[5] Ma Muyang, Ren Pengjie, Lin Yujie, et al. π-Net: A parallel information-sharing network for shared-account cross-domain sequential recommendations[C]//Proc of the 42nd Int ACM SIGIR Conf on Research and Development in Information Retrieval. New York: ACM, 2019: 685−694

[6] Sun Wenchao, Ma Muyang, Ren Pengjie, et al. Parallel split-join networks for shared account cross-domain sequential

recommendations[J]. IEEE Transactions on Knowledge and Data Engineering, 2023, 35(4): 4106−4123

[7] Chen Chen, Guo Jie, Song Bin. Dual attention transfer in session-based recommendation with multi-dimensional integration[C]//Proc of the 44th Int ACM SIGIR Conf on Research and Development in Information Retrieval. New York: ACM, 2021: 869−878

[8] Li Pan, Jiang Zhichao, Que Maofei, et al. Dual attentive sequential learning for cross-domain click-through rate prediction[C]//Proc of the 27th ACM SIGKDD Conf on Knowledge Discovery and Data Mining. New York: ACM, 2021: 3172−3180

[9] Guo Lei, Tang Li, Chen Tong, et al. DA-GCN: A domain-aware attentive graph convolution network for shared-account cross-domain sequential recommendation[C]//Proc of the 30th Int Joint Conf on Artificial Intelligence. Amsterdam: Elsevier, 2021: 2483−2489

[10] Guo Lei, Zhang Jinyu, Tang Li, et al. Time interval-enhanced graph neural network for shared-account cross-domain sequential recommendation[J]. IEEE Transactions on Neural Networks and Learning Systems, 2024, 35(3): 4002−4016

[11] Chen Tong, Yin Hongzhi, Zheng Yujia, et al. Learning elastic embeddings for customizing on-device recommenders[C]//Proc of the 27th ACM SIGKDD Conf on Knowledge Discovery and Data Mining. New York: ACM, 2021: 138−147

[12] Miao Hang, Li Anchen, Yang Bo. Meta-path enhanced lightweight graph neural network for social recommendation[C]//Proc of the 27th Int Conf on Database Systems for Advanced Applications. Berlin: Springer, 2022: 134−149

[13] Wang Ruiqin, Lou Jungang, Jiang Yunliang. LightGCAN: A lightweight graph convolutional attention network for user preference modeling and personalized recommendation[J]. Expert Systems with Applications, 2023, 232: 120741

[14] Zhang Jinyu, Duan Huichuan, Guo Lei, et al. Towards lightweight cross-domain sequential recommendation via external attention-enhanced graph convolution network[C]//Proc of the 28th Int Conf on Database Systems for Advanced Applications. Berlin: Springer, 2023: 205−220

[15] Cheng Chen, Yang Haiqin, Michael R T L, et al. Where you like to go next: Successive point-of-interest recommendation[C]//Proc of the 23rd Int Joint Conf on Artificial Intelligence. Amsterdam: Elsevier, 2013: 2605−2611

[16] He Ruining, Kang Wangcheng, McAuley J. Translation-based recommendation[C]//Proc of the 11th ACM Conf on Recommender Systems. New York: ACM, 2017: 161−169

[17] Song Wei, Qiao Yangyang. Research on the recommender algorithm based on weighted sequential patterns[J]. Journal of Computer Engineering and Science, 2015, 37(7): 1399−1404 (in Chinese)
(宋威, 乔阳阳. 基于加权序列模式的推荐算法研究[J]. 计算机工程与科学, 2015, 37(7): 1399−1404)

[18] Xing Chunxiao, Gao Fengrong, Zhan Sinan, et al. A collaborative filtering recommendation algorithm incorporated with user interest change[J]. Journal of Computer Research and Development, 2007, 44(2): 296−301 (in Chinese)
(邢春晓, 高凤荣, 战思南, 等. 适应用户兴趣变化的协同过滤推荐算法[J]. 计算机研究与发展, 2007, 44(2): 296−301)

[19] Zheng Yujia, Liu Siyi, Li Zekun, et al. DGTN: dual-channel graph transition network for session-based recommendation[C]//Proc of the 2020 Int Conf on Data Mining Workshops. Piscataway, NJ: IEEE, 2020: 236−242

[20] Qiu Ruihong, Huang Zi, Yin Hongzhi, et al. Contrastive learning for representation degeneration problem in sequential recommendation[C]//Proc of the 15th ACM Int Conf on Web Search and Data Mining. New York: ACM, 2022: 813−823

[21] Zhao Rongmei, Sun Siyu, Yan Fanli, et al. Multi-interest aware sequential recommender system based on contrastive learning[J]. Journal of Computer Research and Development, [2024−03−12]. http://kns.cnki.net/kcms/ detail/11.1777.TP.20240105.1351.010.html (in Chinese)
(赵容梅, 孙思雨, 鄢凡力, 等. 基于对比学习的多兴趣感知序列推荐系统[J/OL]. 计算机研究与发展, [2024−03−12]. http://kns.cnki.net/kcms/ detail/11.1777.TP.20240105.1351.010.html)

[22] Quadrana M, Karatzoglou A, Hidasi B, et al. Personalizing session-based recommendations with hierarchical recurrent neural networks[C]//Proc of the 11th ACM Conf on Recommender Systems. New York: ACM, 2017: 130−137

[23] Wu Chaoyuan, Ahmed A, Beutel A, et al. Recurrent recommender networks[C]//Proc of the 10th ACM Int Conf on Web Search and Data Mining. New York: ACM, 2017: 495−503

[24] Zhang Mengqi, Wu Shu, Yu Xueli, et al. Dynamic graph neural networks for sequential recommendation[J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 35(5): 4741−4753

[25] Zhu Tianyu, Sun Leilei, Chen Guoqing. Graph-based embedding smoothing for sequential recommendation[J]. IEEE Transactions on Knowledge and Data Engineering, 2021, 35(1): 496−508

[26] Chen Huiyuan, Lin Yusan, Pan Menghai, et al. Denoising self-attentive sequential recommendation[C]//Proc of the 16th ACM Conf on Recommender Systems. New York: ACM, 2022: 92−101

[27] Xie Jiayi, Chen Zhenzhong. Hierarchical transformer with spatio-temporal context aggregation for next point-of-interest recommendation[J]. ACM Transactions on Information Systems, 2023, 42(2): 1−30

[28] Lin Zhuoyi, Zang Sheng, Wang Rundong, et al. Attention over self-attention: Intention-aware re-ranking with dynamic transformer encoders for recommendation[J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 35(8): 7782−7795

[29] Xie Xu, Sun Fei, Liu Zhaoyang, et al. Contrastive learning for sequential recommendation[C]//Proc of the 38th Int Conf on Data Engineering. Piscataway, NJ: IEEE, 2022: 1259−1273

[30] Ma Muyang, Ren Pengjie, Chen Zhumin, et al. Mixed information flow for cross-domain sequential recommendations[J]. ACM Transactions on Knowledge Discovery from Data, 2022, 16(4): 1−32

[31] Guo Lei, Li Qiuju, Liu Fang'ai, et al. Shared-account cross-domain sequential recommendation with self-attention network[J]. Journal of Computer Research and Development, 2021, 58((11):): 2524−2537 (in Chinese)
(郭磊, 李秋菊, 刘方爱, 等. 基于自注意力网络的共享账户跨域序列推荐[J]. 计算机研究与发展, 2021, 58(11): 2524−2537)

[32] Ren Hao, Liu Baisong, Sun Jinyang, et al. A time and relation-aware graph collaborative filtering for cross-domain sequential recommendation[J]. Journal of Computer Research and Development, 2023, 60(1): 112−124 (in Chinese)
(任豪, 刘柏嵩, 孙金杨, 等. 基于时间和关系感知的图协同过滤跨

域序列推荐[J]. 计算机研究与发展, 2023, 60(1): 112-124)
[33] Liu Siyi, Gao Chen, Chen Yihong, et al. Learnable embedding sizes for recommender systems[C]//Proc of the 9th Int Conf on Learning Representations, Virtual: OpenReview. net, 2021: 1-17
[34] He Xiangnan, Deng Kuan, Wang Xiang, et al. LightGCN: Simplifying and powering graph convolution network for recommendation[C]//Proc of the 43rd Int ACM SIGIR Conf on Research and Development in Information Retrieval. New York: ACM, 2020: 639-648
[35] Guo Menghao, Liu Zhengning, Mu Taijiang, et al. Beyond self-attention: external attention using two linear layers for visual tasks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(5): 5436-5447
[36] Mei M J, Zuber C, Khazaeni Y. A lightweight transformer for next-item product recommendation[C]//Proc of the 16th ACM Conf on Recommender Systems. New York: ACM, 2022: 546-549
[37] Zhuang Fuzhen, Zhou Yingmin, Ying Haochao, et al. Sequential recommendation via cross-domain novelty seeking trait mining[J]. Journal of Computer Science and Technology, 2020, 35: 305-319
[38] Fu Wenjing, Peng Zhaohui, Wang Senzhang, et al. Deeply fusing reviews and contents for cold start users in cross-domain recommendation systems[C]//Proc of the AAAI Conf on Artificial Intelligence. Menlo Park, CA: AAAI, 2019: 94-101
[39] Guo Lei, Zhang Jinyu, Chen Tong, et al. Reinforcement learning-enhanced shared-account cross-domain sequential recommendation[J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 35(7): 7397-7411
[40] Li Xunkai, Guo Ronghui, Chen Jianwen, et al. Neural graph collaborative filtering[C]//Proc of the 42nd Int ACM SIGIR Conf on Research and Development in Information Retrieval. New York: ACM, 2019: 165-174
[41] Kang Wangcheng, McAuley J. Self-attentive sequential recommendation[C]//Proc of the Int Conf on Data Mining. Piscataway, NJ: IEEE, 2018: 197-206
[42] Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks[C]//Proc of the 30th Int Conf on Artificial Intelligence and Statistics. Palermo, Sicily, Italy: JMLR, 2010: 249-256
[43] Kingma D P, Ba J. Adam: A method for stochastic optimization[J]. arXiv preprint, arXiv: 1412.6980, 2014



Zhang Jinyu, born in 1996. PhD candidate. Student member of CCF. His main research interests include sequential recommendation and lightweight recommendation.
张劲羽, 1996 年生. 博士研究生. CCF 学生会员. 主要研究方向为序列推荐、轻量推荐.



Ma Chenxi, born in 1998. Master candidate. Her main research interest includes cross-domain sequential recommendation.
马晨曦, 1998 年生. 硕士研究生. 主要研究方向为跨域序列推荐.



Li Chao, born in 1984. PhD, associate professor. Senior member of CCF. His main research interests include graph data mining and recommendation algorithm.
李 超, 1984 年生. 博士, 副教授. CCF 高级会员. 主要研究方向为图数据挖掘、推荐算法.



Zhao Zhongying, born in 1983. PhD, professor. Distinguished member of CCF. Her main research interests include graph data mining and recommendation system.
赵中英, 1983 年生. 博士, 教授. CCF 杰出会员. 主要研究方向为图数据挖掘、推荐系统.