

利用伪重叠判定机制的多层循环 GCN 跨域推荐*

钱忠胜, 王亚惠, 俞情媛, 范赋宇, 付庭峰



(江西财经大学 计算机与人工智能学院, 江西 南昌 330013)

通信作者: 钱忠胜, E-mail: changesme@163.com

摘要: 跨域推荐 (cross-domain recommendation, CDR) 通过将密集评分辅助域中的用户-项目评分模式迁移到稀疏评分目标域中的评分数据集, 以缓解冷启动现象, 近年来得到广泛研究. 多数 CDR 算法所采用的基于单域推荐的聚类方法未有效利用重叠信息, 无法充分适应跨域推荐, 导致聚类结果不准确. 在跨域推荐中, 图卷积网络方法 (graph convolution network, GCN) 可充分利用节点间的关联, 提高推荐的准确性. 然而, 基于 GCN 的跨域推荐往往使用静态图学习节点嵌入, 忽视了用户的偏好会随推荐场景发生变化的情况, 导致模型在面对不同的推荐任务时表现不佳, 无法有效缓解数据稀疏性. 基于此, 提出一种利用伪重叠判定机制的多层循环 GCN 跨域推荐模型. 首先, 在社区聚类算法 Louvain 的基础上充分运用重叠数据, 设计一个伪重叠判定机制, 据此挖掘用户的信任关系以及相似用户社区, 从而提高聚类算法在跨域推荐中的适应能力及其准确性. 其次, 提出一个包含嵌入学习模块和图学习模块的多层循环 GCN, 学习动态的域共享特征、域特有特征以及动态图结构, 并通过两模块的循环增强, 获取最新用户偏好, 从而缓解数据稀疏问题. 最后, 采用多层感知器 (multi-layer perceptron, MLP) 对用户-项目交互建模, 得到预测评分, 通过与 12 种相关模型在 4 组数据域上的对比结果发现, 所提方法是高效的, 在 *MRR*、*NDCG*、*HR* 指标上分别平均提高 5.47%、3.44%、2.38%.

关键词: 跨域推荐; 伪重叠判定机制; 图卷积网络; 社区聚类; 推荐系统

中图法分类号: TP309

中文引用格式: 钱忠胜, 王亚惠, 俞情媛, 范赋宇, 付庭峰. 利用伪重叠判定机制的多层循环 GCN 跨域推荐. 软件学报. <http://www.jos.org.cn/1000-9825/7289.htm>

英文引用格式: Qian ZS, Wang YH, Yu QY, Fan FY, Fu TF. Multi-layer Recurrent GCN Cross-domain Recommendation with Pseudo-overlap Detection Mechanism. Ruan Jian Xue Bao/Journal of Software (in Chinese). <http://www.jos.org.cn/1000-9825/7289.htm>

Multi-layer Recurrent GCN Cross-domain Recommendation with Pseudo-overlap Detection Mechanism

QIAN Zhong-Sheng, WANG Ya-Hui, YU Qing-Yuan, FAN Fu-Yu, FU Ting-Feng

(School of Computer and Artificial Intelligence, Jiangxi University of Finance & Economic, Nanchang 330013, China)

Abstract: Cross-domain recommendation (CDR) alleviates the cold start problem by transferring the user-item rating patterns from a dataset in a dense rating auxiliary domain to one in a sparse rating target domain, and has been widely studied in recent years. The clustering methods based on single-domain recommendation adopted by most CDR algorithms fail to effectively utilize overlapping information and sufficiently adapt to CDR, resulting in inaccurate clustering results. In CDR, graph convolution network (GCN) methods can fully utilize the associations between nodes to improve recommendation accuracy. However, GCN-based CDR often employs static graph learning for node embedding, ignoring the fact that user preferences may change with different recommendation scenarios, which causes poor model performance across different recommendation tasks and ineffective mitigation of data sparsity. To this end, a multi-layer recurrent GCN CDR model based on a pseudo-overlap detection mechanism is proposed. Firstly, by fully leveraging overlapping data based on the community clustering algorithm Louvain, a pseudo-overlap detection mechanism is designed to mine user trust relationships

* 基金项目: 国家自然科学基金 (62262025); 江西省自然科学基金重点项目 (20224ACB202012)

收稿时间: 2024-04-27; 修改时间: 2024-07-01; 采用时间: 2024-09-02; jos 在线出版时间: 2025-02-26

and similar user communities, thereby enhancing the adaptability and accuracy of clustering algorithms in CDR. Secondly, a multi-layer recurrent GCN consisting of an embedding learning module and a graph learning module is proposed to learn dynamic domain-shared features, domain-specific features, and dynamic graph structures. By conducting iterative enhancement of the two modules, the latest user preferences are obtained to alleviate data sparsity. Finally, a multi-layer perceptron (MLP) is employed to model user-item interactions and obtain predicted ratings. Comparative results with 12 related models across four groups of data domains demonstrate the effectiveness of the proposed method, with average improvements of 5.47%, 3.44%, and 2.38% in *MRR*, *NDCG*, and *HR* metrics respectively.

Key words: cross-domain recommendation; pseudo-overlap detection mechanism; graph convolution network (GCN); community clustering; recommender system

推荐系统 (recommender system, RS) 的基本思想是从大量的候选项目中选择用户可能感兴趣的项目并推荐给用户. 然而, 在大多数实际应用场景中存在数据稀疏问题, 这使得推荐系统无法为用户生成最恰当表示, 同时新用户/项目也引起冷启动现象. 为缓解数据稀疏以及冷启动问题, 一些基于协同过滤 (collaborative filtering, CF) 的推荐方法开始引入迁移学习技术^[1], 从数据较密集的领域中提取共享知识, 用于改进目标领域的推荐性能, 称为跨域推荐 (cross-domain recommendation, CDR) 方法.

CDR 方法通过引入辅助域可有效缓解推荐中冷启动问题^[2-5], 学习不同域之间的重叠用户/项目的潜在表征, 同时结合域间共享信息作为连接辅助域和目标域的桥梁. 一些模型只考虑对单个用户的表示进行建模, 使用不同方法 (如循环生成网络、元网络) 学习个性化的双向传递函数. 然而, 这些方法未能从用户群体出发, 且未兼顾到跨域信息的差异性, 其适用范围不广. 人类行为易受其他个体的影响, 特别是朋友和具有相似偏好的用户. 根据协同过滤思想^[6], 可从相似用户中提取共同特征并应用于改进和学习个性化建模表示. 但随着 CDR 数据量的增加, 数据的稀疏性和高维性现象不断显现, 模型通过将相似用户/项目进行聚类, 帮助发现相似兴趣的用户/项目群组, 可有效缓解该问题. 在推荐系统中虽然协同过滤在聚类算法中运用最为广泛^[7], 但在跨域推荐领域, 协同过滤无法高效处理不同领域之间的复杂关系. 同时, 基于聚类的 CDR 模型往往面向用户完全重叠场景而很少针对部分用户重叠的推荐场景进行改进, 使得聚类效率低下, 处理速度受到影响. 以源域和目标域分别为电影和书籍为例, 假设用户 u_1, u_2, u_3 均在电影领域与多处相同项目交互而在书籍领域的交互信息较少, 可看作 u_1, u_2, u_3 具有相似的电影偏好, 那么 u_2 在书籍领域的交互信息对于 u_1, u_3 也具有一定的参考价值. 而在用户部分重叠场景下, 比如, 用户 u_1 在书籍领域无交互信息, 而 u_3 在电影领域无交互信息, 则他们的重叠用户 u_2 就成为连接 u_1, u_3 的桥梁. 这种情况下无法直接套用常规基于单域推荐的方法进行聚类, 有必要对聚类方法进行改进以利用重叠用户信息, 为后续特征学习提供更准确的聚类结果.

因此, 设计能针对部分用户重叠推荐场景的聚类方法对提高跨域推荐的适应能力及其准确性至关重要.

在跨域推荐中, 一些传统非图的 CDR 方法难以捕捉复杂的用户-项目交互, 只能通过一跳邻居交互信息学习用户表征. 为获取高阶信息, 人们利用图神经网络 (graph neural network, GNN)^[8] 的信息提取能力, 不仅可运用一跳邻居信息, 还可使用通过图结构间接连接的多跳邻居信息, 获得更高质量嵌入特征的高阶关系. 然而, 现有大多数 GNN 方法主要关注节点特征信息的利用, 而在跨域场景中, 节点却可能缺乏明显的特征. 为提高跨域节点嵌入的准确性, 研究人员开始探索在不使用节点特征信息的情况下, 如何更有效地学习跨域节点嵌入^[9]. 考虑到不同域及不同项目对用户目标域兴趣偏好具有不同的影响, 一些基于图卷积网络方法 (graph convolution network, GCN) 的 CDR 模型主要通过学习域共享信息和域特有信息来获取高阶的跨域节点嵌入^[10]. 在使用 GCN 前, 需要通过图的形式构建用户-项目节点关系作为 GCN 的初始输入, 但事实上, 随着节点嵌入的学习, 固定的图构建过程会忽略不同域间节点关系的变化.

因此, 在基于 GCN 的 CDR 模型中增加循环结构并使图随着节点嵌入的学习而变化, 可提高节点特征的鲁棒性和数据的利用率, 从而缓解数据稀疏问题.

基于上述两个方面的问题, 本文提出一种利用伪重叠判定机制的多层循环 GCN 跨域推荐模型 PO-CDRec, 主要工作与贡献如下.

(1) 在已有的用户-项目交互基础上, 针对跨域推荐中部分用户重叠场景, 创新性地改进社区聚类算法^[11], 提出

伪重叠判定机制, 使得伪重叠用户社区面积最大化, 充分利用重叠信息并整合不同域信息学习域共享信息, 捕获用户-用户信任关系, 进而构建源域与目标域一体化的用户网络, 提高聚类算法对跨域推荐的适应性, 从而解决聚类结果准确性不高问题。

(2) 分别从单域和联合域网络出发, 提出了包含嵌入学习模块和图学习模块的多层循环 GCN 算法, 并利用伪重叠判定机制得到的节点相似矩阵, 在动态图增强中学习域共享信息和域特有信息以获取高阶的跨域节点嵌入, 学习用户最新偏好, 提高数据利用率, 从而缓解数据稀疏问题。

(3) 通过 MLP 对用户-项目交互建模, 得到预测交互概率, 在 4 组关于 CDR 数据域上的对比实验结果表明, PO-CDRec 在 *MRR*、*NDCG* 以及 *HR* 这 3 个评价指标上均有所提升。此外, 还进行了消融实验分析, 验证了 PO-CDRec 模型组件的有效性。

1 相关工作

跨域推荐旨在利用多个不同领域的数据来提供更准确、个性化更强的推荐。在跨域推荐领域, GCN 扮演着关键角色, 可通过学习不同领域数据之间的图结构关系提升跨域推荐的个性化程度, 但也存在图结构数据过于复杂的问题, 通过引入聚类方法可在图中形成更清晰的群组结构, 增强数据的可解释性。

聚类方法通过距离度量方式划分数据以发现节点的相似性和规律性, 以更好地理解节点间的交互行为和关系, 从而提供更精确和个性化的推荐。Afoudi 等人^[12]将自组织映射方法与协同过滤相结合, 利用自组织映射对节点聚类, 发现数据潜在的非线性模式, 更好地捕捉用户和物品间的关系, 但易造成数据过拟合, 从而降低推荐系统的泛化性和鲁棒性。Siet 等人^[13]在利用 Transformer 机制获取项目嵌入后, 以 K-Means 对项目嵌入分组, 并将聚类信息与预测的评分相结合以提高个性化推荐的多样性。但该模型对初始值的选择较敏感, 不同初始值可能会导致聚类结果有所差别, 使结果不稳定。Chang 等人^[14]通过基于度量学习将松散的项目序列重构为紧密的项目-项目兴趣图, 将长期用户行为中不同类型的偏好整合到图聚类中, 并在兴趣图上进行图卷积传播以提取兴趣。但基于度量学习的兴趣图易受非结构化数据影响。这类方法通过引入社交网络提高推荐准确性, 然而采用直接单一的社交关系无法适应不同领域的用户偏好。以上模型均基于单域推荐提出, 接下来介绍跨域推荐的聚类算法。Kang 等人^[15]提出一种半监督框架 SSCDR, 分别学习源域和目标域度量空间的用户和项目潜在向量, 同时训练映射函数捕获两个度量空间的跨域关系。根据冷启动用户在源域的邻接节点推断其潜在向量为冷启动用户推荐目标域项目。Xu 等人^[16]提出 HCCDR 模型将软聚类方法运用于识别每个目标用户的相似用户, 并同时使用从相似用户学习到的共同特征和用户的个人特征来学习用户的个性化偏好传递函数。软聚类算法需计算每个数据点属于每个聚类的概率, 这在处理大规模数据集时可能影响算法效率。Liu 等人^[17]针对隐私保护问题提出多域推荐模型 FPPDM, 通过局部组件学习各域用户/项目偏好分布后, 利用全局聚合组件获取跨域组合用户特征, 并提出紧致性联合聚类方法对特征相似的用户聚类以获得更紧凑的用户表示, 但多域推荐中各域数据稀疏性不同会导致各域特征分布差异大, 使得 FPPDM 无法有效适用该聚类方法。然而, K-Means 聚类和软聚类等传统算法往往适用于数据维度较低、样本规模适中的场景, 而无法有效处理跨域推荐中大规模、高维度的数据集。

图神经网络方法通过构建共享图表示用户之间的关系, 学习图表示捕提高阶非线性节点之间的依赖关系。Wang 等人^[18]首次将图神经网络应用于跨域社交推荐, 并提出神经社交协作排名 NSCR 方法, 利用属性感知的深度协同过滤模型在信息导向领域学习用户-项目交互。然而 NSCR 未能充分利用 GNN 聚合不同领域的用户信息, 导致在源域到目标域的信息扩散力度较弱。Cui 等人^[19]提出一个异构图框架 HeroGRAPH, 该框架利用每个域内的信息进行域内建模, 并且运用共享图上的图卷积操作进行跨域建模。以类似的方式, Xu 等人^[20]提出基于关系扩展的跨域推荐模型 ReCDR。不同于 HeroGRAPH 框架构造共享图的方式, ReCDR 利用图嵌入模型来为两个域中的所有节点生成预训练的节点嵌入, 使得具有较高相似性的节点被连接从而设计共享图。然而以上模型的用户行为均在单个域中处理, 这种处理方式使跨域信息合并时很不融洽。Li 等人^[21]提出一种嵌入式内容和异构网络模型 ECHCDR, 创造性地结合了对抗学习算法。该模型训练生成器和鉴别器来学习用户和项目表示, 基于所学表示的内

积进行域内和域间推荐. 尽管以上模型有一定程度上的改进但均未同时考虑两个域间的交互. Cao 等人^[22]提出一种新的方法 CDRIB, 利用信息瓶颈的原则, 设计了两个 IB 正则化器以同时对域间/域内用户-项目交互进行建模, 从而兼顾了两个域的交互使之可共同去偏置化. 这些正则化器鼓励对域共享信息进行编码, 使域共享信息能直接在两个域中进行推荐. 然而, 虽然以上模型均是为优化基于图结构的推荐而设计的, 但它们仍是从传统 GCN 的角度来设计的, 未能很好地考虑跨域推荐中的静态化偏好问题.

通过上面的分析, 当前运用于跨域推荐的聚类算法未能很好地利用重叠数据, 且很少考虑与大规模 CDR 数据集的适配性, 导致聚类算法速度不高, 且无法保证聚类结果准确性. 同时跨域推荐中 GCN 方法使用静态图学习节点嵌入, 无法有效处理数据稀疏问题.

基于此, 本文提出一种利用伪重叠判定机制的多层循环 GCN 跨域推荐模型 PO-CDRec. 该模型充分运用重叠数据对 Louvain 社区聚类算法进行改进, 设计一个伪重叠判定机制, 挖掘用户的信任关系以及相似用户社区, 从而提高聚类算法在跨域推荐中的准确率和效率; 此外, 该模型还利用一个包含嵌入学习模块和图学习模块的多层循环 GCN, 学习动态的域共享特征、域特有特征以及动态图结构, 并通过两模块的循环增强, 获取用户最新偏好, 从而缓解数据稀疏问题, 提高跨域数据利用率. 在得到用户与项目的组合特征向量后, 运用采用 MLP 对用户-项目交互建模, 获取最终预测交互概率.

2 利用伪重叠判定机制的多层循环 GCN 跨域推荐模型框架

问题定义: 假定源域和目标域中分别存在一组用户及其对应的交互项目, 两组用户部分重叠, 项目不重叠, 分别用集合 $U^S = U^O \cup \{u_1, u_2, \dots, u_{|U^S - U^O|}\}$, $U^T = U^O \cup \{u_1, u_2, \dots, u_{|U^T - U^O|}\}$, $I^S = \{i_1, i_2, \dots, i_{|I^S|}\}$ 和 $I^T = \{i_1, i_2, \dots, i_{|I^T|}\}$ 表示, 其中 $|U^S|$ 、 $|U^T|$ 、 $|I^S|$ 和 $|I^T|$ 分别为源域用户数、目标域用户数、源域项目数和目标域项目数. 重叠用户集用 $U^O = \{u_1, u_2, \dots, u_{|U^O|}\}$ 表示, 其中 $|U^O|$ 为重叠用户数. 对于重叠用户 $u \in U^O$, 他在源域与目标域中均存在交互的项目 I_u^S 和 I_u^T . 这样, 跨域推荐任务被定义为, 模型利用重叠用户的交互信息链接源域与目标域用户, 并分别学习不同域用户的域共享特征和域特有特征, 最终为两域中不重叠用户做推荐.

本文模型 PO-CDRec 利用伪重叠判定机制对用户聚类以获取用户相似度矩阵, 同时使用动态循环图结构的多层 GCN 学习域特有和域共享节点嵌入, 最终预测源域及目标域得分概率. 下面给出 PO-CDRec 的整体框架, 如后文图 1 所示, 主要包含如下几个模块.

1) 伪重叠判定机制模块, 获取用户相似性. 利用重叠用户对联合域用户进行聚类时, 以社区聚类算法取代传统跨域中的聚类算法, 我们设计了伪重叠用户判断机制, 改进 Louvain 算法用于跨域推荐, 对源域和目标域的用户进行聚类, 生成用户社交网络, 得到用户信任关系.

2) 多层循环图卷积网络模块, 挖掘高阶节点嵌入. 利用 GCN 聚合节点嵌入时, 以动态循环图结构代替传统图卷积中的固定图学习. 通过这种方式聚合嵌入可充分利用不同域数据信息, 从域内/域间角度考虑权重关系, 增强模型的高阶连通性, 并利用门控网络聚合最终的节点嵌入表示.

3) 模型预测模块, 生成推荐列表. 采用 MLP 对用户-项目交互建模, 获取源域与目标域的预测得分概率. 联合源域与目标域的损失函数优化总损失函数, 提升模型的泛化能力, 达到更优的推荐效果.

为便于阐述, 对文中一些主要符号进行说明, 如后文表 1 所示.

2.1 基于伪重叠 Louvain 算法的社区聚类

跨域推荐根据用户是否与源域/目标域项目同时交互细分为重叠用户与非重叠用户(冷启动用户), 利用传统聚类方法的 CDR 模型往往对两域用户分别聚合生成用户群体, 这忽视了不同领域用户的差异性, 同时也无法适用于大规模 CDR 数据集. 相对于传统聚类方法, 社区聚类将用户划分为不同社区, 能够更好地处理复杂网络中的用户节点间的关系, 从而挖掘具有相似兴趣的用户群体, 并可有效适用于大规模数据集. Louvain 算法^[23]作为一种广泛流行的社区聚类方法具备可扩展性和高效性, 其使用模块度作为相似性度量的局部优化策略, 逐步将节点聚集成社区, 并将社区视为单个节点继续进行迭代, 以形成模块度最大化的社区聚类, 大大减少了计算复杂度. 为使其

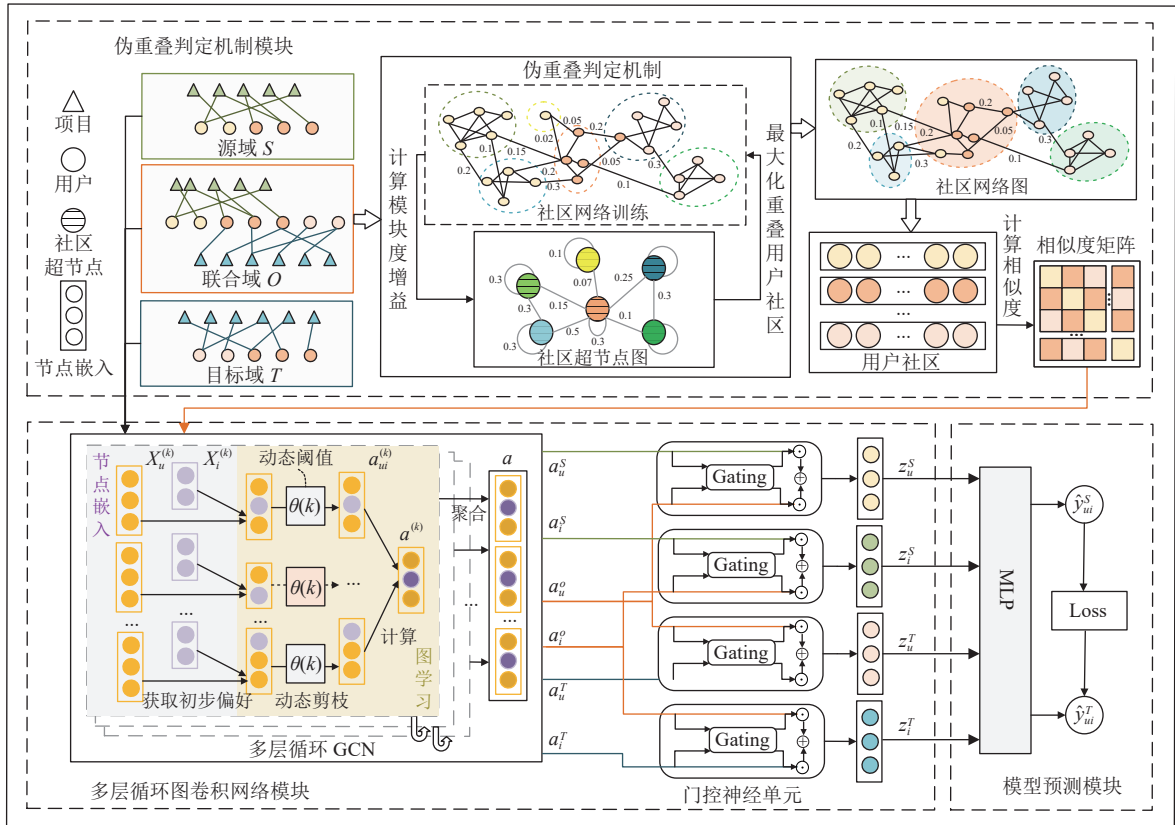


图 1 利用伪重叠判定机制的多层循环 GCN 跨域推荐框架

表 1 主要符号及含义

符号	含义
U^S, U^T	源域/目标域用户集
I^S, I^T	源域/目标域项目集
Q	模块度
ΔQ	模块度增益
m	社区网络所有边权重和
C	社区集
\sum_c^{in}	社区 c 内部边权重和
\sum_c^{tot}	与社区 c 中节点相连边的权重和
U^o	重叠用户集
λ	重叠用户权重
β	共同邻居权重
K	预定义的嵌入传播深度
X_u, X_i	用户/项目嵌入矩阵
z^S, z^T	源域/目标域组合特征向量
\hat{y}^S, \hat{y}^T	源域/目标域得分概率

更适配于部分用户重叠场景, 我们结合重叠用户交互信息对其做出相应改进, 提出伪重叠判定机制, 便于跨域推荐.

给定两个域, 分别为源域 S 和目标域 T , 其中用户集部分重叠, 项目集不重叠, 源域/目标域用户集、源域/目标域项目集分别记为 U^S/U^T 和 I^S/I^T . 源域-目标域联合相似度矩阵 W , 如公式 (1) 所示.

$$W = \begin{bmatrix} W^S & 0 \\ 0 & W^T \end{bmatrix} \quad (1)$$

其中, W^S 表示源域矩阵, W^T 表示目标域矩阵.

为更好地实现不同域信息的传递, 我们从用户群体出发, 对用户关系建模, 挖掘相似度高的用户群体, 通过联合源域与目标域的用户-项目评分矩阵计算用户间的相似度, 如公式 (2) 所示.

$$w_{uv} = \frac{\sum_{i \in I_{uv}} (r_{u,i} - \bar{r}_u) \cdot (r_{v,i} - \bar{r}_v)}{\sqrt{\sum_{i \in I_{uv}} (r_{u,i} - \bar{r}_u)^2} \sqrt{\sum_{i \in I_{uv}} (r_{v,i} - \bar{r}_v)^2}} \quad (2)$$

其中, w_{uv} 表示用户 u 和用户 v 的相似度, $r_{u,i}$ 是用户 u 对项目 i 的评分, $r_{v,i}$ 是用户 v 对项目 i 的评分, \bar{r}_u 和 \bar{r}_v 分别为用户 u 、 v 交互项目的平均评分, I_{uv} 表示用户 u 、 v 均交互的项目集. 可利用同样方法获取项目间的相似度.

在跨域推荐场景中, 需考虑不同域间的关系, 为更好地整合不同域间的信息, 本文引入并改进 Louvain 算法, 以获取不同域间的用户关系和项目关系, 从而实现域间社区聚类. 模块度作为相似度社区划分的度量方法, 取值范围为 $[-0.5, 1]$, 其计算如公式 (3) 所示.

$$Q = \sum_{c \in C} \left[\frac{\sum_c^{in}}{2m} - \left(\frac{\sum_c^{tot}}{2m} \right)^2 \right] \quad (3)$$

其中, Q 表示模块度, C 表示社区集, m 是社区网络所有边权重和, \sum_c^{in} 表示社区 c 内部边权重和, \sum_c^{tot} 表示与社区 c 中节点相连边的权重和.

在聚类过程中, Louvain 算法采用模块度增益来动态调整聚类结果, 通过将节点移入使模块度增益达到最大的邻居社区, 此时总体社区网络模块度达到最大, 相似度社区划分也呈现最佳效果, 如公式 (4) 所示.

$$\Delta Q_{u \rightarrow c} = \left[\frac{\sum_c^{in} + \sum_{v \in c} w_{uv}}{2m} - \left(\frac{\sum_c^{tot} + w_u}{2m} \right)^2 \right] - \left[\frac{\sum_c^{in}}{2m} - \left(\frac{\sum_c^{tot}}{2m} \right)^2 - \left(\frac{w_u}{2m} \right)^2 \right] = \frac{\sum_{v \in c} w_{uv}}{2m} - \frac{\sum_c^{tot} * w_u}{2m^2} \quad (4)$$

其中, $\sum_{v \in c} w_{uv}$ 表示社区 c 内部与节点 u 关联边的权重和, w_u 表示所有与 u 关联边的权重和, 即 u 的权重.

然而, Louvain 算法无法很好地适应不同域间的聚类, 我们考虑使用重叠用户作为联合域的桥梁, 对该算法加以改进. 为充分利用域间关系并发挥重叠用户的作用, 本文提出伪重叠判定机制, 其原理是通过相似度判断机制扩大重叠用户面积. 例如, 假设用户 u 为重叠用户, 而非重叠用户 v 与之存在高度相似性, 则可认为 v 也近似为重叠用户, 以此方式扩大重叠用户面积, 最终使得重叠用户面积最大化, 如公式 (5) 所示.

$$\Delta Q_{u \rightarrow c} = \begin{cases} \frac{\sum_{v \in c} w_{uv}}{2m} - \frac{\sum_c^{tot} * w_u}{2m^2} + \lambda \sum_{v \in c} w_{uv}, & v \in U^o \\ \frac{\sum_{v \in c} w_{uv}}{2m} - \frac{\sum_c^{tot} * w_u}{2m^2}, & v \notin U^o \end{cases} \quad (5)$$

其中, U^o 表示重叠用户集, λ 为重叠用户权重.

通过公式 (5), 形成重叠面积最大化的伪重叠用户社区, 充分利用更新后的伪重叠用户信息, 使得面向单个网络的 Louvain 算法可通过伪重叠用户社区连接双域网络. 基于伪重叠判定机制的聚类方法具体实施如下.

在初始阶段, 每个用户被视为一个社区. 先随机选择初始节点, 然后对每个节点分别进行如下操作, 并多次迭代.

步骤 1. 计算所有节点模块度增益. 对于节点 u , 计算将 u 从目前所在社区移入每个邻居节点所在社区 c 所产生模块度增益 ΔQ . 首先, 源域、目标域中的用户向量被随机初始化为同一维度, 将这些域中的用户向量视为用户联合域 $ST \in \mathbb{R}^{d \times (|S|+|T|)}$, 并将其作为 Louvain 算法的输入; 其次, 利用域间用户相似度并结合 Louvain 聚类算法获得初步社区结果; 最后, 由于社区聚类涉及多个数据域, 信息量过大, 为减少聚类偏差, 采用模块度增益来动态调整聚类结果.

在对 u 的所有邻居节点计算完成后, 将 u 移入能获得最大 ΔQ 的社区, 若 ΔQ 不为正, 则节点 u 保留在原社区中.

步骤 2. 基于社区聚类结果构建超图. 每个社区被压缩成一个带有自环边的超节点, 该自环边的权重为社区内部权重和. 超节点间边的权重为两两社区中节点间的边权重和.

重复步骤 1、2, 直至社区网络模块度达到最优, 得到完整的社区网络, 可获取最终的社区集. 在社交网络中, 共同邻居的数量可反映用户之间的兴趣相似程度. 若两个用户有很多共同的朋友, 则他们很可能有相似的兴趣和行为. 故我们使用社区网络图中的共同邻居重新衡量用户间的相似度. 同一社区中的用户 u 和 v 的相似度计算, 如公式 (6) 所示.

$$w'_{uv} = w_{uv} + \beta \sum_{k \in c} (w_{uk} + w_{vk}), (k \neq u \cap v) \cap (w_{uk} \cap w_{vk}) \quad (6)$$

其中, w'_{uv} 表示经过伪相似判定后重新获取的用户相似度, β 表示共同邻居权重.

计算用户 u 的所有相似用户节点, 获得基于 u 的相似度矩阵 W' , 即联合域相似度矩阵, 它由源域相似度矩阵与目标域相似度矩阵组成. 同时相似度矩阵 W' 也是加权邻接矩阵, 可作为多层循环 GCN 节点嵌入学习模块的输入. 伪重叠判定机制的流程图如图 2 所示.

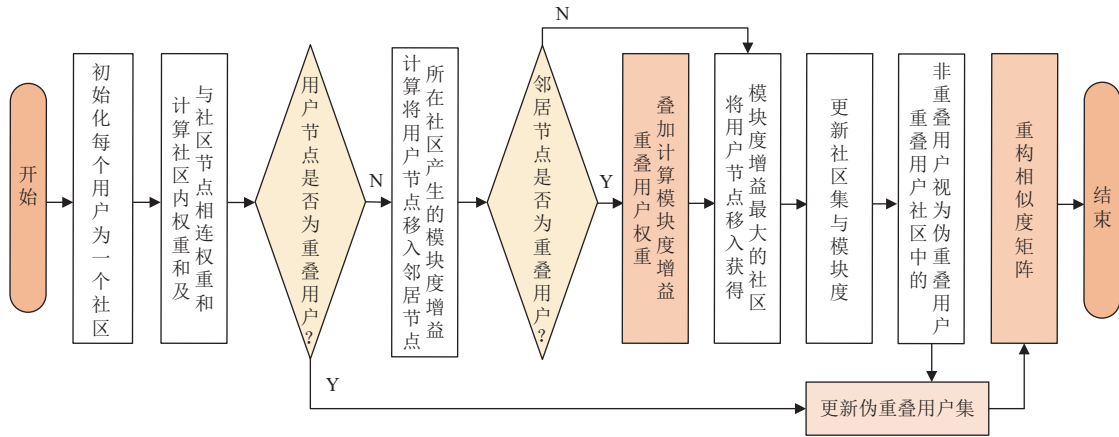


图 2 伪重叠判定机制流程图

2.2 基于多层循环 GCN 的偏好获取

跨域场景中节点大多缺乏明显的特征, 故为提高跨域节点嵌入的准确性, 考虑不使用用户-项目特征信息捕获节点嵌入. 我们提出多层循环 GCN 学习跨域节点嵌入, 包含节点嵌入模块和图学习模块. 其中, 节点嵌入模块涉及域共享嵌入层和两域特有嵌入层, 分别生成用于跨域传输的共享嵌入和两域特定嵌入; 图学习模块将用户、项目嵌入作为输入, 得到用户-项目偏好.

在第 2.1 节利用伪重叠判定机制得到所有用户、项目相似度矩阵后, 将其作为多层循环 GCN 节点嵌入模块的初始输入. 该模块利用 GCN 分别学习源域、目标域和联合域节点嵌入, 通过聚合低阶信息, 得到用户、项目域共享和两域特有嵌入, 并利用门控神经网络得到用户组合嵌入和项目组合嵌入. 多层循环 GCN 图如图 3 所示.

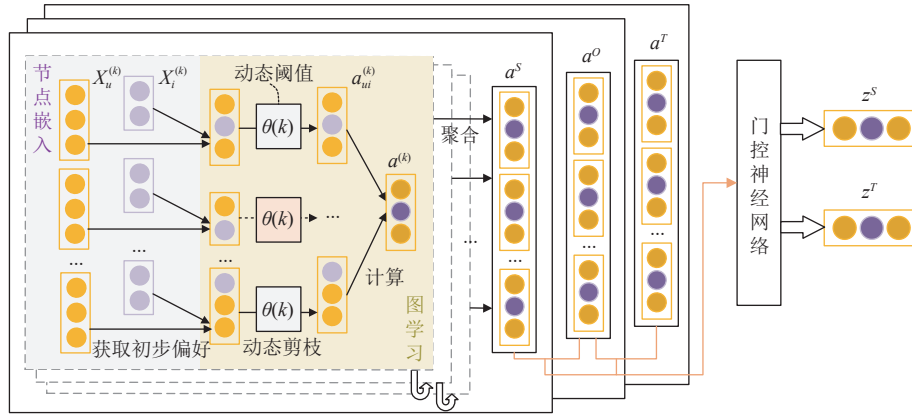


图3 多层循环 GCN 图

嵌入层的输入是邻接矩阵和节点特征矩阵,通过对节点属性和图结构进行编码生成节点嵌入.在嵌入层中采用多层 GCN,可有效地学习基于邻域聚合的节点嵌入.在初始阶段,每个用户(或项目)节点初始化为特征向量 X_u^0 (或 X_i^0),此时图结构分别为用户、项目相似度图.节点嵌入计算如公式(7)所示.

$$\begin{cases} X_u^{(k+1)} = \varpi(D_u^{-\frac{1}{2}} \hat{W}_u D_u^{-\frac{1}{2}} X_u^{(k)} F_u^{(k)}) \\ X_i^{(k+1)} = \varpi(D_i^{-\frac{1}{2}} \hat{W}_i D_i^{-\frac{1}{2}} X_i^{(k)} F_i^{(k)}) \end{cases} \quad (7)$$

其中, $X_u^{(k)}$ 和 $X_i^{(k)}$ 分别是第 k 层得到的用户、项目嵌入矩阵, ϖ 表示激活函数, \hat{W}_u 和 \hat{W}_i 分别表示带自连接的用户、项目邻接矩阵, $\hat{W} = W + I$ (I 为单位矩阵), D_u 和 D_i 为用户、项目节点的度矩阵,表示与之相似度不为零的邻居节点数, $F_u^{(k)}$ 和 $F_i^{(k)}$ 分别是第 k 层的用户、项目权重矩阵.

嵌入模块得到的用户、项目节点嵌入作为图学习模块的输入,该模块在获取新的权重矩阵后,将其作为下一层节点嵌入模块的再次输入,多次迭代操作.这样,利用图学习模块将用户、项目表示可从潜在空间映射到相似度空间,获取用户-项目偏好,学习方式如公式(8)所示.

$$a_{ui}^{(k+1)} = \sigma \left(\frac{\langle X_u^{(k+1)} \times F_1, X_i^{(k+1)} \times F_2 \rangle}{|X_u^{(k+1)} \times F_1| |X_i^{(k+1)} \times F_2|} \right) \quad (8)$$

其中, $a_{ui}^{(k+1)}$ 表示第 $(k+1)$ 层得到的用户-项目偏好, $\sigma(\cdot)$ 是 *sigmoid* 函数,将相似度归一化在 $(0, 1)$ 区间内, F_1 和 F_2 是两个可训练的矩阵, $\langle \cdot \rangle$ 表示向量内积操作.

当输入数据维数较大时,公式(8)所得的偏好矩阵会过于稠密,同时结合前文中使用的伪重叠判定机制,模块计算时间和内存使用量成倍增长,而对于大规模数据集,内存消耗会较高,很可能导致计算设备内存不足.故无法直接运用于 GCN,需对得到的用户-项目偏好矩阵进行稀疏化处理.

许多模型在稀疏化处理过程中采用幅度剪枝(magnitude pruning, MaP)方法^[24],但 MaP 侧重于局部信息,忽略全局结构和权重间的关联.另外,相似度较小的数据剪枝也可能导致模型性能下降.为应对此问题,我们使用自适应方法,基于相似度的幅度设定相应阈值,在训练过程中动态调整,从而使偏好矩阵不过于稠密.对于每位用户对应项目的偏好与上一轮的阈值共同构成该层的预测阈值.

为更充分地利用图数据,避免过度剪枝可能导致的性能下降,在训练过程中,使用动态调整机制^[25]对不同用户或不同情境自适应调整,以更灵活地应对局部信息和全局结构的变化.同时,为平衡剪枝过程中各轮次之间的关系,并确保每一轮的剪枝阈值更加独立和合理,设置权重参数,使得每一轮的剪枝比例逐渐降低,从而减少前几轮对后续轮次的影响.这里通过自适应幅度剪枝进行动态调整,动态阈值及剪枝如公式(9)、(10)所示.

$$\theta_{ui}(k) = \alpha a_{ui}^{(k-1)} + (1 - \alpha) \theta_{ui}(k-1) \quad (9)$$

$$a_{ui}^{(k)} = \begin{cases} a_{ui}^{(k)}, & a_{ui}^{(k)} \geq \theta_{ui}(k) \\ 0, & a_{ui}^{(k)} < \theta_{ui}(k) \end{cases} \quad (10)$$

其中, $\theta_{ui}(k)$ 是动态阈值, α 为权重参数, 决定历史数据对当前的影响程度.

每层均对用户 u 所有项目偏好计算, 得到该层的偏好向量, 经过 K 步聚合传播后, 获得用户最终嵌入向量, 如公式 (11) 所示.

$$a_u = a_u^{(0)} \parallel \dots \parallel a_u^{(k)} \parallel \dots \parallel a_u^{(K)} \quad (11)$$

其中, \parallel 表示连接操作, $a_u^{(k)}$ 是用户 u 的第 k 层偏好向量, $k=0, 1, 2, \dots, K$, a_u 为用户 u 的最终嵌入向量.

根据上述分析, 输入联合域信息并通过循环 GCN 获得用户 u 的域共享嵌入向量 a_u^o , 输入源域/目标域信息获取对应的源域/目标域特有嵌入向量 a_u^s/a_u^t ; 同样得到项目 i 的域共享嵌入向量和域特有嵌入向量 $a_i^o, a_i^s/a_i^t$.

在平衡联合域共性与单域特性的前提下, 为有效地组合用户 u 的共享嵌入向量与特有嵌入向量, 我们分别使用共享部分参数的门控神经元生成源域与目标域的组合特征向量 z_u^s, z_u^t (项目 i 在源域与目标域的组合特征向量 z_i^s, z_i^t 也由同样方法生成), 其计算如公式 (12) 所示.

$$\begin{cases} z_u^s = \sigma(Vt_u^o a_u^o + Vt_u^s a_u^s) \odot a_u^o + [1 - \sigma(Vt_u^o a_u^o + Vt_u^s a_u^s)] \odot a_u^s, \\ z_u^t = \sigma(Vt_u^o a_u^o + Vt_u^t a_u^t) \odot a_u^o + [1 - \sigma(Vt_u^o a_u^o + Vt_u^t a_u^t)] \odot a_u^t \end{cases} \quad (12)$$

其中, $Vt_u^o, Vt_u^s/Vt_u^t$ 分别是门控单元的域共享权重矩阵和域特有权重矩阵.

2.3 模型预测过程

在第 2.2 节获取用户与项目的组合特征向量后, 为更好地学习用户与项目表示之间的交互, 我们采用 MLP 模型对用户-项目交互建模, 以源域为例, 如公式 (13)、(14) 所示.

$$\begin{cases} \varphi^{(1)} = \omega^{(1)}(G^{(1)} \begin{bmatrix} z_u^s \\ z_i^s \end{bmatrix} + h^{(1)}) \\ \vdots \\ \varphi^{(L)} = \omega^{(L)}(G^{(L)}\varphi^{(L-1)} + h^{(L)}) \end{cases} \quad (13)$$

$$\hat{y}_{ui}^s = f(\varphi^{(L)}) \quad (14)$$

其中, $\varphi^{(k)}$ 表示第 k 层的输出结果 ($1 \leq k \leq L$), $G^{(k)}$ 和 $h^{(k)}$ 分别为第 k 层可训练矩阵和偏置项, $\omega^{(k)}(\cdot)$ 表示第 k 层激活函数, $f(\cdot)$ 表示将 $\varphi^{(L)}$ 映射为得分概率 \hat{y}_{ui}^s 的预测函数.

目标域的评分预测 \hat{y}_{ui}^t 也由同样方法得到. 利用伪重叠判定机制的多层循环 GCN 跨域推荐算法, 利用伪重叠判定机制获取节点相似度矩阵后, 输入至多层循环 GCN 以有效学习并组合节点特征, 最终利用 MLP 学习源域和目标域的得分概率, 如算法 1 所示.

算法 1. 利用伪重叠判定机制的多层循环 GCN 跨域推荐过程.

输入: 源域/目标域用户集 U^s/U^t , 源域/目标域项目集 I^s/I^t , 源域/目标域评分矩阵 R^s/R^t ;

输出: 社区集 C , 模块度 Q , 源域/目标域的预测得分概率 \hat{y}_{ui} (即 $\hat{y}_{ui}^s/\hat{y}_{ui}^t$).

Begin

1. $C \leftarrow \{\{u\}\}, \forall u \in U$; // 初始化每位用户为一个社区
 2. $\sum_c^{in} \leftarrow \sum w_{uv}, u \in c \cap v \in c$; // 计算社区 c 内部边权重和
 3. $\sum_c^{tot} \leftarrow \sum w_{uv}, u \in c \cup v \in c$; // 计算与社区 c 中节点相连边的权重和
- /* 伪重叠判定机制模块 */
4. **For** $u \in U$ // 找到用户 u 的最优社区
 5. **For** $v \in U \cap v \in c \cap v \neq u \cap w_{uv} \neq 0$
-

$$6. \quad \Delta Q_{u \rightarrow c} = \begin{cases} \frac{\sum_{v \in c} W_{uv}}{2m} - \frac{\sum_c^{tot} * W_u}{2m^2} + \lambda \sum_{v \in c} W_{uv}, & v \in U^o; \\ \frac{\sum_{v \in c} W_{uv}}{2m} - \frac{\sum_c^{tot} * W_u}{2m^2}, & v \notin U^o \end{cases};$$

// 计算将 u 移入所有邻居节点 v 所在社区 c 所产生的模块度增益, 见公式 (5)

7. **End For**

8. **If** $\max \Delta Q_{u \rightarrow \hat{c}}$ **then** // 将 u 移入使其获得最大模块度增益的社区 \hat{c}

$$9. \quad \sum_{\hat{c}}^{in} \leftarrow \sum_{\hat{c}}^{in} + \sum_{v \in \hat{c}} W_{uv}, \quad \sum_{\hat{c}}^{tot} \leftarrow \sum_{\hat{c}}^{tot} + W_u;$$

$$10. \quad \sum_c^{in} \leftarrow \sum_c^{in} - \sum_{v \in c} W_{uv}, \quad \sum_c^{tot} \leftarrow \sum_c^{tot} - W_u;$$

$$11. \quad \hat{c} \leftarrow \hat{c} \cup \{u\}, \quad c \leftarrow c - \{u\};$$

12. **End If**

13. **End For**

14. **For** $c \in C$ // 更新社区集并计算模块度

$$15. \quad Q \leftarrow Q + \frac{\sum_c^{in}}{2m} - \left(\frac{\sum_c^{tot}}{2m} \right)^2;$$

$$16. \quad C' \leftarrow C' \cup \{c\};$$

17. **End For**

18. **Print** C' and Q ; // 输出社区集和模块度

19. **For** $u \in c$ // 重构相似度矩阵, 见公式 (6)

20. **For** $v \in c \cap v \neq u$

$$21. \quad w'_{uv} = w_{uv} + \beta \sum_{k \in c} (w_{uk} + w_{vk}), (k \neq u \cap v) \cap (w_{uk} \cap w_{vk});$$

22. **End For**

23. **End For**

/* 多层循环 GCN 模块 */

24. **For** $u \in U, i \in I$ // 分别聚合用户/项目节点与相应邻居节点

$$25. \quad X^{(k+1)} = \omega (D^{-\frac{1}{2}} \hat{W} D^{-\frac{1}{2}} X^{(k)} F^{(k)}); // 生成节点嵌入, 见公式 (7)$$

$$26. \quad a_{ui}^{(k+1)} = \sigma \left(\frac{\langle x_u^{(k+1)} \times F_1, x_i^{(k+1)} \times F_2 \rangle}{|x_u^{(k+1)} \times F_1| |x_i^{(k+1)} \times F_2|} \right); // 生成用户-项目偏好权重, 见公式 (8)$$

$$27. \quad \theta_{ui}(k) = \alpha a_{ui}^{(k-1)} + (1 - \alpha) \theta_{ui}(k-1) // 阈值动态调整, 见公式 (9)$$

28. **If** $a_{ui}^{(k)} < \theta_{ui}(k)$ **then** // 对偏好权重进行动态剪枝, 见公式 (10)

$$29. \quad a_{ui}^{(k)} = 0;$$

30. **End If**

31. **End For**

$$32. \quad x = x^{(0)} \parallel \dots \parallel x^{(k)} \parallel \dots \parallel x^{(K-1)}; // 聚合节点的 K 层嵌入, 见公式 (11)$$

$$33. \quad z^s = \sigma (V^O x^O + V^S x^S) \odot x^O + [1 - \sigma (V^O x^O + V^S x^S)] \odot x^S;$$

$$34. \quad z^T = \sigma (V^O x^O + V^T x^T) \odot x^O + [1 - \sigma (V^O x^O + V^T x^T)] \odot x^T;$$

// 生成源域/目标域节点组合特征, 见公式 (12)

/* 模型预测模块 */

$$35. \quad \varphi^{(L)} = \omega^{(L)} (G^{(L)} \varphi^{(L-1)} + h^{(L)}), \hat{y}_{ui} = f(\varphi^{(L)}); // 获得源域/目标域的预测得分概率, 见公式 (13)、(14)$$

36. **Output** \hat{y}_{ui} ; // 输出源域/目标域的最终预测得分概率

End

在算法 1 中, 第 1–23 行是伪重叠判定机制模块, 通过伪重叠判定机制对用户聚类得到社区集和模块度, 并以此重新构建相似度矩阵; 第 24–34 行是多层循环 GCN 模块, 其中第 24–32 行利用多层动态 GCN 获得域特有和域共享节点嵌入, 第 32–34 行通过 4 个门控神经元聚合得到源域和目标域的用户、项目组合特征向量; 第 35–36 行是模型预测模块, 使用 MLP 得到不同域用户的最终预测项目得分概率。

源域和目标域损失函数被定义为预测得分和实际得分的交叉熵, 如公式 (15) 所示。

$$\begin{aligned} L_S &= - \sum_{(u,i) \in \Omega^+ \cup \Omega^-} (y_{ui}^S \log \hat{y}_{ui}^S + (1 - y_{ui}^S) \log (1 - \hat{y}_{ui}^S)), \\ L_T &= - \sum_{(u,i) \in \Omega^+ \cup \Omega^-} (y_{ui}^T \log \hat{y}_{ui}^T + (1 - y_{ui}^T) \log (1 - \hat{y}_{ui}^T)) \end{aligned} \quad (15)$$

其中, y_{ui}^s / y_{ui}^t 是已知的用户-项目交互, Ω^+ 是观察到的交互集, Ω^- 是从未观察到的交互集中随机抽样的一定数量负实例, 以防止过拟合。

为进一步提高两域的推荐性能, 我们将前文中提出的损失函数相结合, 得到联合优化的总损失函数, 如公式 (16) 所示。

$$L = \lambda_1 L_S + \lambda_2 L_T + \lambda_3 \|\Theta\|_2^2 \quad (16)$$

注意在公式 (16) 中, 考虑到两域中数据集稀疏性不一致问题, 本文通过调节权重 λ_1 和 λ_2 平衡源域和目标域的样本, 以控制样本平衡性。比如, 若目标域的数据更为稀疏, 则可设定较大的 λ_2 , 以增加目标域数据对总损失的影响。 λ_3 则为控制正则化的权重, Θ 表示网络参数, 两者共同构成正则化项以防止模型过拟合。

3 实验设计与分析

3.1 问题提出

为阐明模型的有效性及其优势, 本文选择经典数据集 Amazon 的 4 组数据域进行综合实验对比与分析, 主要回答下面 3 个问题。

RQ1: 与经典的、较新的相关模型对比, 本文模型有何优势? 效果如何?

针对此问题, 第 3.4.1 节设置了对比实验, 分别将本文模型 PO-CDRec 与 12 种相关模型作对比。实验结果表明, 在 *MRR*, *NDCG*, *HR* 这 3 个评价指标上, 模型 PO-CDRec 均优于其他对比模型。可知, 所提模型在评估用户对未知项目评分方面有很大优势, 能捕获用户最新偏好并提升推荐性能。

RQ2: 本文模型的不同构件对模型自身的性能影响如何? 是否有存在的必要性?

针对此问题, 在第 3.4.2 节设置了消融实验, 根据伪重叠判定机制、多层循环 GCN 这两个构件组成 3 种变体模型, 分析这些构件对模型推荐性能的影响。实验结果表明, 这 3 种变体模型的推荐效果均在不同程度上劣于所提模型 PO-CDRec, 说明这两个构件在解决 CDR 问题上具有一定效果。

RQ3: 超参数如何影响本文模型的性能?

针对此问题, 在第 3.4.3 节设置了参数敏感度实验, 重点选择传播深度和重叠用户权重展开分析, 发现它们在不同的数据集上使模型达到最优值时的取值有所不同, 并分析了其潜在的原因, 以便更好地优化模型。

3.2 数据集与评价指标

3.2.1 数据集

本文根据多数 CDR 研究选择的跨域数据集和预处理设置来构建 CDR 场景, 在大规模公共亚马逊数据集上进行实验。该 Amazon (来源 <http://jmcauley.ucsd.edu/data/amazon/>) 数据集由 24 个不相交的项目域组成。针对用户部分重叠、项目不重叠的推荐场景, 本文选取部分数据域构成 4 组源域-目标域, 分别是: Music-Movie、Phone-Elec、Cloth-Sport、Game-Video。为了将目标域中的项目推荐给只存在源域中的用户 (冷启动用户), 故在数据预处理中, 本文过滤掉在域中少于 10 个交互的项目和少于 5 个交互的用户, 使用户-项目能够从其源域中学习具有代表性的嵌入。同时随机选择约 20% 的重叠用户作为冷启动用户进行测试和验证, 剩余的用户用于训练。具体信息如表 2 所示, 其中, 训练集这一列中数据表示从数据域中选取用于模型训练的用户-项目交互数。

表2 数据集统计信息

数据域	用户	项目	训练集	重叠用户	冷启动用户
Music	50 841	43 858	674 233		1 893
Movie	87 875	38 643	1 127 424	15 081	1 885
Phone	27 519	9 481	148 271		2 049
Elec	107 984	40 460	821 301	16 337	2 042
Cloth	41 829	17 943	187 880		990
Sport	27 328	12 655	163 291	7 857	981
Game	25 025	12 319	155 036		226
Video	19 457	8 751	156 091	1 737	217

3.2.2 评价指标

为验证模型 PO-CDRec 的推荐性能是否有所提高, 选取常用的平均倒数秩 (mean reciprocal rank, *MRR*)、归一化折损累计增益 (normalize discounted cumulative gain, *NDCG*)、命中率 (hit rate, *HR*) 这 3 个评价指标对其性能进行评价。

1) 平均倒数秩. 它是准确推荐项目倒数秩的平均值, 当所有秩位置超过 n 时, 倒数秩设为 0, 如公式 (17) 所示。

$$MRR = \frac{1}{n} \sum_{u=1}^n \frac{1}{rank_u} \quad (17)$$

其中, n 表示用户数, $rank_u$ 表示第 u 个用户的排名。

2) 归一化折损累计增益. 它表示准确推荐项目在推荐列表中位置越靠前, 其贡献越大, 反之贡献越小, 如公式 (18) 所示。

$$NDCG = \sum_{u \in U_{test}} \frac{1}{Y_u} \sum_{i=1}^n \frac{2^{t_i} - 1}{\log_2(i+1)} \quad (18)$$

其中, U_{test} 表示用户测试集, Y_u 表示用户真实值, t_i 表示项目 i 在推荐列表中的位置。

3) 命中率. 它强调模型推荐的准确性, 即用户的需求项是否包含在模型的推荐列表中, 如公式 (19) 所示。

$$HR = \frac{1}{s} \sum_{i=1}^s hit(i) \quad (19)$$

其中, s 为样本数, 可理解为用户需求项的数目. $hit(i)$ 用于表示第 i 个项目是否包含在模型推荐的项目列表中, 若是, 则其值为 1; 否则为 0。

3.3 对比模型

因 PO-CDRec 属于 CDR 领域范畴, 故将其与经典的、先进的 CDR 方法进行比较. 与本文模型 PO-CDRec 作对比的 12 种模型包括 3 个单域推荐模型和 9 个跨域推荐模型. 在单域推荐模型中, CML^[26]为推荐领域的经典模型; TCCF^[27]将改进的聚类算法应用在推荐模型中, 与本文模型的伪重叠判定机制有一定联系; NGCF^[28]是基于 GCN 的经典推荐方法, 与本文模型的多层循环 GCN 模块有一定相关性. 在跨域推荐模型中, EMC^[29]和 STAR^[30]均为跨域推荐中的经典模型; SSCDR、CDAML^[31]、HCCDR、FPPDM 为适配跨域推荐分别对聚类方法进行不同方面的改进, 与本文模型的伪重叠判定机制具有相似性, 其中 SSCDR 在 EMC 基础上拓展; PPGN^[32]、I²RCDR^[33]、GLOU^[34]分别运用 GCN 学习跨域推荐中不同角度的嵌入表示, 与本文模型的多层循环 GCN 模块具有一定相关性。

单域推荐: 两个域的所有交互被合并为一个域, 应用以下广泛使用的基于 CF 的方法。

1) CML 在度量学习思想下学习用户和项目表示, 该度量学习思想利用 L2 距离并假设用户表示与交互项目的距离小于尚未交互的项目。

2) TCCF 引入布谷鸟搜索算法改进 K-Means 算法聚合用户, 并利用时间相关系数捕捉用户行为的时间模式。

3) NGCF 是一种典型的基于图神经网络学习用户和项目表示的方法, 它堆叠 3 层 GCN 以探索用户和项目间的高阶连接信息。

跨域推荐: 跨域推荐模型分别为每个域学习不同的用户-项目表示, 下面是几种典型的跨域基线模型.

1) EMCDDR 是第 1 个为冷启动用户提供 CDR 的推荐框架, 其采用 3 种不同的预训练方法, 包括 CML、TCCF 和 NGCF, 分别初始化 EMCDDR 中的用户-项目表示, 称为 EMCDDR(CML)、EMCDDR(TCCF) 和 EMCDDR(NGCF).

2) STAR 设计了一个编码器拓扑结构, 并为所有域引入一个共享的基础网络来传输知识.

3) SSCDDR 对 EMCDDR 进行扩展, 先通过 CML 方法预训练表示, 再利用聚合重叠用户的多跳相邻信息学习跨域映射函数.

4) CDAML 提出一种软聚类的方法, 在更细粒度的聚类级别指导全局共享参数初始化, 并通过对抗学习转移与领域无关的用户偏好以改进目标领域的推荐.

5) HCCDDR 采用软聚类的方法来识别每个目标用户的相似用户, 并同时结合从相似用户中学习到的共同特征和自身特征以获取用户的个性化偏好传递函数.

6) FPPDM 通过局部组件运用各领域评级信息来学习用户/项目偏好分布后, 利用全局聚合组件获取跨域组合用户特征.

7) PPGN 通过共享用户嵌入层中两个单独的 GCN 学习用户-项目表示, 并采用基于 EMCDDR 的方法, 旨在应对用户冷启动推荐.

8) I²RCCR 设计一个关系感知的 GCN, 通过探索不同图间的多跳异构连接, 提取域共享和域特定特征, 并结合引入的门控融合机制, 实现双重目标推荐.

9) GLOU 运用一种自监督分类器, 根据不同领域用户偏好差异将用户分组, 并利用 GCN 对其交互项形成的子图进行学习, 挖掘高阶协作关系.

3.4 实验及其分析

对于各个模型的公共超参数均采用相同值, 包括: 嵌入维度 F 为 128, 批量大小为 1024. 对于基线模型中的特定超参数, 我们使用原始参考模型中的值. 此外, 对于所有 EMCDDR 变体模型, 我们采用与 EMCDDR 中建议的相同 MLP 架构. 对于本文模型 PO-CDRec, 学习率在 $\{0.01, 0.001, 0.001\}$ 集合中调整, GCN 网络层数在 $\{0, 1, 2, 3, 4\}$ 中调整. 本文模型使用 Adam 作为优化器, 而对于基线模型, 保留了文献中使用的原始优化器. 对所有模型, 每个实验运行 5 次, 并在验证集中选择最佳结果.

下面结合 3 种综合实验阐述本文模型的有效性及其优势. 其中, 对比实验 (见第 3.4.1 节) 将本文模型与 12 种经典的、较新的推荐模型进行对比, 为保证实验的公平性, 对比实验均选取各模型中所需较为合理的数据集划分比例, 阐明本文模型的优势, 以回答 RQ1; 消融实验 (见第 3.4.2 节) 阐述本文模型的相关构件对模型推荐性能提高的可行性与高效性, 以回答 RQ2; 参数敏感度实验 (见第 3.4.3 节) 分析主要参数对模型性能方面的影响, 便于实施模型参数优化, 以回答 RQ3.

3.4.1 模型对比分析 (RQ1)

表 3-表 6 给出了本文模型 PO-CDRec 与 12 种对比模型在 4 组数据域下的 3 个评价指标 MRR , $NDCG@N$, $HR@N$ 实验结果值. 实验中测试了 N 为 5 与 10 时的效果. 为便于比较, 利用虚线区分单域模型和跨域模型.

表 3 Music-Movie 数据域中推荐模型性能比较

模型	Music-domain recommendation					Movie-domain recommendation				
	MRR	$NDCG$		HR		MRR	$NDCG$		HR	
		@5	@10	@5	@10		@5	@10	@5	@10
CML	4.19±0.09	3.50±0.11	4.53±0.14	5.50±0.17	8.70±0.27	3.89±0.04	2.91±0.04	3.95±0.03	4.62±0.05	7.87±0.11
TCCF	4.78±0.05	3.67±0.08	4.51±0.20	5.72±0.11	8.82±0.09	4.68±0.12	3.34±0.07	4.62±0.13	5.59±0.03	9.10±0.10
NGCF	4.04±0.04	3.18±0.06	4.15±0.09	5.03±0.08	8.86±0.10	4.08±0.08	3.02±0.06	4.53±0.10	5.24±0.10	9.92±0.16
EMCDDR(CML)	3.80±0.01	3.06±0.01	3.97±0.02	4.78±0.01	7.62±0.04	4.73±0.05	3.68±0.05	4.97±0.06	5.81±0.08	9.82±0.10
EMCDDR(TCCF)	4.41±0.10	3.71±0.18	4.65±0.11	5.41±0.19	8.79±0.18	5.56±0.03	4.31±0.09	5.79±0.08	6.50±0.14	10.10±0.12
EMCDDR(NGCF)	4.36±0.08	3.63±0.04	4.70±0.08	5.61±0.09	8.95±0.12	5.49±0.12	4.55±0.08	6.06±0.14	7.25±0.14	11.89±0.10

表3 Music-Movie 数据域中推荐模型性能比较(续)

模型	Music-domain recommendation					Movie-domain recommendation				
	MRR	NDCG		HR		MRR	NDCG		HR	
		@5	@10	@5	@10		@5	@10	@5	@10
STAR	3.91±0.12	3.12±0.12	4.07±0.13	4.78±0.15	7.71±0.20	3.91±0.18	2.73±0.18	4.00±0.23	4.62±0.25	8.60±0.04
SSCDR	1.95±0.01	1.36±0.01	2.15±0.02	2.15±0.02	3.59±0.04	2.74±0.01	1.99±0.01	2.72±0.01	3.19±0.03	5.47±0.03
CDAML	4.99±0.02	3.74±0.04	6.25±0.16	6.79±0.09	10.07±0.03	6.14±0.16	5.01±0.06	6.12±0.11	8.19±0.17	13.26±0.06
HCCDR	5.49±0.13	4.16±0.12	6.47±0.06	6.65±0.17	10.14±0.06	6.25±0.10	5.27±0.04	6.24±0.08	8.06±0.13	14.09±0.11
FPPDM	6.75±0.13	5.58±0.07	7.62±0.03	9.11±0.23	13.26±0.12	6.47±0.09	5.20±0.04	6.81±0.05	8.82±0.12	14.26±0.01
PPGN	4.18±0.04	3.39±0.06	4.55±0.07	5.32±0.11	8.91±0.16	4.31±0.03	3.19±0.02	4.65±0.04	5.44±0.07	10.00±0.10
I ² RCDR	4.96±0.12	3.87±0.03	6.13±0.05	6.84±0.18	9.48±0.13	5.97±0.05	4.72±0.08	6.25±0.15	8.02±0.20	13.07±0.12
GLOU	6.25±0.17	5.95±0.19	7.46±0.12	9.27±0.06	14.86±0.11	6.59±0.02	5.24±0.05	7.11±0.10	8.84±0.15	14.78±0.12
PO-CDRec	7.13±0.04	6.21±0.03	7.91±0.01	9.31±0.06	14.98±0.13	6.64±0.26	5.81±0.07	7.41±0.13	9.02±0.16	14.89±0.04

表4 Phone-Elec 数据域中推荐模型性能比较

模型	Phone-domain recommendation					Elec-domain recommendation				
	MRR	NDCG		HR		MRR	NDCG		HR	
		@5	@10	@5	@10		@5	@10	@5	@10
CML	5.83±0.10	5.14±0.11	6.41±0.13	7.62±0.13	11.56±0.23	6.24±0.13	5.54±0.17	6.85±0.12	8.34±0.23	12.44±0.10
TCCF	6.97±0.15	6.47±0.12	7.51±0.09	9.06±0.13	11.98±0.19	6.64±0.16	6.37±0.10	7.71±0.13	8.96±0.24	13.97±0.26
NGCF	6.48±0.04	5.82±0.08	7.47±0.12	9.11±0.19	13.62±0.20	8.04±0.07	7.41±0.08	9.09±0.11	11.01±0.17	16.24±0.21
EMCDR(CML)	5.69±0.01	5.06±0.03	6.32±0.02	7.96±0.04	11.89±0.05	7.28±0.01	6.61±0.01	8.17±0.01	9.93±0.02	14.75±0.03
EMCDR(TCCF)	6.46±0.05	5.86±0.04	7.37±0.06	8.95±0.07	13.76±0.03	8.33±0.01	7.85±0.08	9.35±0.02	11.22±0.01	16.54±0.06
EMCDR(NGCF)	6.93±0.04	6.35±0.02	7.73±0.03	9.54±0.03	13.84±0.07	8.77±0.02	8.10±0.02	9.78±0.02	11.76±0.03	17.01±0.03
STAR	4.96±0.22	4.22±0.20	5.41±0.21	6.48±0.23	10.22±0.22	6.67±0.30	5.86±0.32	7.42±0.34	8.92±0.41	13.76±0.47
SSCDR	3.33±0.01	2.63±0.01	3.33±0.01	3.97±0.02	6.14±0.03	4.94±0.01	4.21±0.01	5.50±0.01	6.28±0.02	10.34±0.03
CDAML	7.41±0.02	7.16±0.11	8.05±0.12	10.52±0.01	14.19±0.16	9.34±0.08	9.02±0.21	10.42±0.03	12.67±0.03	18.01±0.06
HCCDR	7.49±0.15	7.57±0.04	8.36±0.20	11.02±0.05	14.77±0.08	9.75±0.13	9.63±0.10	10.48±0.23	12.80±0.06	18.43±0.10
FPPDM	7.47±0.11	8.12±0.04	9.27±0.13	11.38±0.02	16.59±0.05	10.42±0.17	9.66±0.04	10.87±0.14	13.06±0.12	19.78±0.05
PPGN	6.51±0.09	6.04±0.08	7.39±0.16	9.49±0.16	13.60±0.23	8.07±0.03	7.39±0.06	9.16±0.05	10.87±0.10	16.38±0.13
I ² RCDR	7.05±0.09	6.98±0.13	7.86±0.08	10.13±0.12	14.17±0.10	8.95±0.13	8.84±0.09	10.05±0.04	12.12±0.05	17.95±0.09
GLOU	7.84±0.07	8.06±0.15	9.14±0.15	11.80±0.15	17.67±0.17	11.02±0.17	10.14±0.23	11.38±0.09	14.54±0.15	20.21±0.12
PO-CDRec	8.08±0.05	8.21±0.17	9.32±0.25	12.46±0.11	17.90±0.04	11.28±0.08	10.21±0.18	11.63±0.15	14.67±0.03	20.43±0.16

表5 Cloth-Sport 数据域中推荐模型性能比较

模型	Cloth-domain recommendation					Sport-domain recommendation				
	MRR	NDCG		HR		MRR	NDCG		HR	
		@5	@10	@5	@10		@5	@10	@5	@10
CML	3.73±0.16	3.19±0.19	3.92±0.14	4.69±0.21	6.97±0.11	3.26±0.14	2.65±0.15	3.29±0.16	3.82±0.17	5.82±0.20
TCCF	3.18±0.05	2.36±0.06	3.35±0.10	3.99±0.16	6.79±0.14	3.31±0.11	2.59±0.14	3.16±0.09	3.74±0.22	5.89±0.25
NGCF	3.24±0.08	2.50±0.08	3.48±0.13	4.02±0.12	7.07±0.30	3.42±0.06	2.75±0.04	3.63±0.07	4.43±0.10	7.22±0.11
EMCDR(CML)	2.70±0.02	1.96±0.01	2.72±0.01	3.04±0.02	5.42±0.06	2.91±0.02	2.18±0.03	2.99±0.04	3.46±0.02	6.02±0.06
EMCDR(TCCF)	4.36±0.03	3.94±0.05	4.56±0.07	5.48±0.05	7.36±0.13	3.63±0.07	2.72±0.08	3.71±0.04	4.56±0.07	7.48±0.13
EMCDR(NGCF)	4.89±0.06	4.08±0.06	5.17±0.08	5.55±0.10	7.91±0.15	4.07±0.09	3.24±0.10	4.03±0.12	4.94±0.15	7.41±0.16
STAR	3.89±0.13	3.33±0.11	4.15±0.16	5.20±0.17	7.76±0.27	3.00±0.16	2.30±0.18	3.19±0.18	3.73±0.22	6.49±0.22
SSCDR	2.96±0.01	2.27±0.01	3.06±0.04	3.64±0.03	6.12±0.05	3.46±0.01	2.66±0.01	3.75±0.02	4.18±0.02	7.27±0.02
CDAML	4.43±0.11	3.98±0.13	4.59±0.08	7.21±0.05	7.21±0.05	3.26±0.03	2.73±0.03	3.72±0.02	4.39±0.03	7.51±0.02
HCCDR	5.56±0.07	4.76±0.05	4.98±0.07	7.72±0.16	7.53±0.11	3.68±0.10	2.84±0.14	4.02±0.12	4.86±0.23	7.87±0.22
FPPDM	6.14±0.13	5.27±0.02	5.13±0.10	7.56±0.17	7.23±0.08	4.64±0.16	3.28±0.02	4.69±0.08	7.84±0.20	11.18±0.06
PPGN	3.34±0.09	2.58±0.10	3.57±0.04	4.05±0.16	7.10±0.12	3.30±0.10	2.54±0.08	3.58±0.11	4.52±0.10	7.36±0.20

表 5 Cloth-Sport 数据域中推荐模型性能比较 (续)

模型	Cloth-domain recommendation					Sport-domain recommendation				
	MRR	NDCG		HR		MRR	NDCG		HR	
		@5	@10	@5	@10		@5	@10		
I ² RCDR	4.83±0.10	4.28±0.08	5.03±0.12	5.77±0.06	8.11±0.16	3.68±0.03	3.14±0.03	3.84±0.04	4.60±0.06	7.18±0.12
GLOU	6.19±0.12	5.55±0.23	5.81±0.28	8.30±0.19	11.19±0.29	4.59±0.12	4.89±0.10	5.62±0.09	7.90±0.14	11.04±0.20
PO-CDRec	6.28±0.05	5.64±0.12	5.84±0.17	8.61±0.12	11.89±0.27	5.18±0.18	5.06±0.13	5.87±0.07	8.03±0.21	11.42±0.28

表 6 Game-Video 数据域中推荐模型性能比较

模型	Game-domain recommendation					Video-domain recommendation				
	MRR	NDCG		HR		MRR	NDCG		HR	
		@5	@10	@5	@10		@5	@10		
CML	1.70±0.08	1.10±0.08	1.44±0.09	1.74±0.09	2.82±0.18	1.46±0.14	0.71±0.10	1.30±0.08	1.25±0.13	3.07±0.10
TCCF	2.42±0.14	1.57±0.03	1.97±0.15	1.77±0.24	3.72±0.24	2.68±0.21	1.86±0.22	2.65±0.24	2.85±0.21	4.58±0.19
NGCF	2.94±0.07	2.07±0.02	2.73±0.09	3.14±0.16	5.14±0.22	3.72±0.05	2.86±0.03	3.87±0.10	4.25±0.18	7.41±0.18
EMCDR(CML)	2.12±0.01	1.35±0.03	1.82±0.04	2.17±0.07	3.62±0.09	3.15±0.01	2.57±0.01	3.23±0.02	4.25±0.04	6.34±0.05
EMCDR(TCCF)	2.42±0.11	1.59±0.14	2.34±0.08	2.56±0.30	4.63±0.13	4.15±0.06	3.36±0.04	4.29±0.05	4.87±0.10	7.95±0.32
EMCDR(NGCF)	2.47±0.18	1.86±0.15	2.44±0.14	3.26±0.20	5.07±0.17	3.98±0.01	3.21±0.02	4.29±0.02	5.09±0.08	8.43±0.04
STAR	2.17±0.09	1.61±0.04	2.17±0.16	2.53±0.12	4.27±0.40	2.62±0.10	1.87±0.01	2.62±0.11	3.00±0.14	5.30±0.18
SSCDR	1.98±0.03	1.27±0.01	1.59±0.03	1.96±0.03	3.48±0.06	2.55±0.03	1.86±0.02	2.61±0.02	3.14±0.06	5.51±0.08
CDAML	2.78±0.04	1.83±0.02	2.88±0.06	3.53±0.05	5.84±0.13	3.73±0.04	2.61±0.05	3.79±0.06	4.20±0.01	7.46±0.13
HCCDR	3.12±0.12	2.17±0.14	2.78±0.16	3.45±0.31	5.60±0.21	3.79±0.09	2.95±0.12	3.93±0.14	4.39±0.17	7.48±0.32
FPPDM	4.43±0.15	3.23±0.10	3.68±0.06	5.47±0.04	7.16±0.02	3.95±0.14	3.36±0.03	4.57±0.16	5.69±0.08	9.35±0.05
PPGN	2.77±0.06	1.90±0.09	2.84±0.10	3.07±0.14	5.98±0.32	3.90±0.11	3.04±0.07	4.06±0.16	4.53±0.09	7.68±0.32
I ² RCDR	2.52±0.05	1.97±0.02	2.58±0.07	3.40±0.05	5.36±0.09	3.97±0.07	3.15±0.07	4.41±0.08	4.95±0.05	8.85±0.11
GLOU	4.52±0.16	3.71±0.11	4.58±0.13	5.44±0.21	8.61±0.30	4.75±0.13	4.65±0.01	5.49±0.14	7.41±0.21	12.11±0.23
PO-CDRec	4.87±0.12	3.87±0.07	4.92±0.04	5.61±0.09	9.17±0.01	5.02±0.15	4.68±0.02	5.68±0.16	7.52±0.24	12.18±0.13

表 3–表 6 展示了所有模型在各数据域性能, 其中, 本文模型 PO-CDRec 在数据集 Amazon 的数据域 Music-Movie、Phone-Elec、Cloth-Sport 中达到最优, 在数据域 Game-Video 中整体表现良好. 下面给出更详细分析.

3.4.1.1 与单域推荐模型对比

对比模型 TCCF 和 NGCF 分别是基于改进聚类方法和图神经网络的单域推荐模型, 性能在许多方面都优于 CML. 例如, 在表 6 中 Game-Video 的 CDR 场景下, TCCF 模型的 MRR 值比 CML 模型平均提升 37.64%. 这证实了捕获高阶邻居信息以及改进用户聚类算法对于用户和项目表示的学习具有一定帮助.

在多数推荐场景下, NGCF 模型性能略优于 TCCF. 例如, 在表 5 中 Cloth-Sport 的 CDR 场景下, NGCF 模型的 NDCG@5、HR@5 值比 TCCF 模型分别平均提升 5.7% 和 8.17%. 这表明在单域推荐中, 使用 GCN 捕获高阶邻居信息相较于改进用户聚类算法对推荐性能的提升会产生更大影响.

相较于 3 个对应的单域推荐模型, 跨域框架 EMCDR 可更好地适配于跨域推荐场景, 其实验效果取得显著的改进. 例如, 在表 4 中 Phone-Ele 的 CDR 场景下, EMCDR(CML)、EMCDR(TCCF)、EMCDR(NGCF) 的 HR@10 值相较于其对应的单域模型 CML、TCCF、NGCF 分别平均提升 9.23%、14.24%、3.06%. 这主要因为传统单域方法未考虑跨域差异, 难以辨别冷启动用户的信息传输.

表 3–表 6 的实验结果表明, 这些方法虽然在一定程度上有效, 但均仅基于单域方式进行推荐, 模型的泛化能力不足, 很难适应不同领域的推荐任务, 同时也初步验证本文模型中对聚类方法以及 GCN 模块改进的有效性.

3.4.1.2 与跨域推荐模型对比

相比于对应的单域推荐模型 CML、TCCF 和 NGCF 而言, 基于 EMCDR 框架的跨域推荐模型 EMCDR (CML)、EMCDR(TCCF) 和 EMCDR(NGCF) 性能更为突出. 这验证了跨域框架 EMCDR 在缓解冷启动问题上具

有一定的可行性. 由表 3–表 6 可知, EMCDR(TCCF) 和 EMCDR(NGCF) 在所有数据域中表现均高于 EMCDR(CML). 例如, 在 Game 数据域上, EMCDR(TCCF) 和 EMCDR(NGCF) 的 MRR 值比 EMCDR(CML) 分别提高 12.40%、14.17%. 这表明基于改进的聚类算法和图神经网络策略可较好地适应跨域推荐.

表 4 和表 6 的 Phone 数据域、Game 数据域的结果中, EMCDR 的方法改进非常有限, 甚至在某些情况下比它们对应的单域推荐模型更差, 其中在 Phone 数据域对应的 MRR 值平均下降 1.07%, 这表明简单的跨域映射不足以描述跨域用户表示的复杂映射关系.

对比模型 STAR 和 PPGN 相对于单域推荐模型而言, 性能未有明显提升, 是因为这些方法侧重于信息传递来增强重叠用户推荐, 并且易给冷启动用户生成有偏见的表示.

为提高映射函数的能力, SSCDR 考虑用户-项目多跳邻居信息以增强 EMCDR(CML) 的函数映射效果. 表 3–表 6 中除了表 5 的 Cloth-Sport 数据域外, SSCDR 在其他数据域的指标反而普遍不如 EMCDR(CML), 其中在 Music 数据域中, SSCDR 模型的 MRR 值比 EMCDR(CML) 模型降低 48.68%. 这是因为 SSCDR 对初始化的预训练表示较为敏感.

CDAML、HCCDR、FPPDM 均采用聚类方法生成用户共同特征, 模型性能在多数情况下高于利用 EMCDR 框架的基线模型. 其中在表 5 中 Cloth-Sport 的 CDR 场景下, CDAML、HCCDR、FPPDM 的 $NDCG@5$ 值比 EMCDR(TCCF) 模型分别提升 0.69%、10.73%、15.2%, 这表明在 CDR 中聚类方法是有效的.

\hat{I}^2 RCDR 与 GLOU 都是基于 GCN 的跨域推荐模型, 与基于 EMCDR 框架的基线模型相比, 这两个模型性能均得到一定改进, 其中在表 3 中 Music-Movie 的 CDR 场景下, \hat{I}^2 RCDR 模型的 $NDCG@5$ 值比 EMCDR(NGCF) 模型平均提升 4.9%, 而 GLOU 模型的 $NDCG@5$ 值比 EMCDR(NGCF) 模型平均提升 26.09%. GLOU 模型实验效果优于 \hat{I}^2 RCDR、CDAML、HCCDR、FPPDM 模型, 在对比模型中表现最佳, 主要是因为 GLOU 模型在使用 GCN 提取节点特征前会根据偏好差异对用户分组, 这表明将聚类方法与 GCN 相结合是可行的.

表 3–表 6 的实验结果表明, 基于改进的聚类方法和图神经网络策略在 CDR 中均取得一定效果. 与这些模型不同的是, 所提模型 PO-CDRec 改进聚类算法的同时还考虑了用户和项目的高阶嵌入关系, 使其具备更强优势.

综上分析, 相较于之前的 CDR 模型, 本文模型 PO-CDRec 在所有数据域的评价指标中均获得较好的表现. 与对比模型中表现最佳的 GLOU 模型相比, 在 Music-Movie、Phone-Elec、Cloth-Sport 和 Game-Video 这 4 组数据域上, 所提模型 PO-CDRec 的 MRR 值最少提升 6.55%、2.65%、6.41%、6.29%; $NDCG@N$ 值最少提升 4.85%、1.26%、2.39%、2.38%; $HR@N$ 值最少提升 0.77%、1.19%、2.61%、2.25%.

从对比结果来看, 本文模型 PO-CDRec 在 Music-Movie、Phone-Elec 数据域中的表现更好. 相对于 Cloth-Sport 与 Game-Video 这两组数据域, 模型的 $HR@N$ 值在 Music-Movie 数据域中分别最少提升 9.27% 和 28.51%, 而在 Phone-Elec 数据域中分别最少提升 38.69% 和 44.29%.

本文模型 PO-CDRec 在 Music-Movie 与 Phone-Elec 数据域上表现更出色的原因在于, 这两组数据域中的重叠用户信息更多, 模型适应性更强. 一方面, 模型 PO-CDRec 中的伪重叠判定机制模块运用了基于重叠用户改进的社区聚类算法, 可适用于大规模数据集, 能有效处理不同领域的复杂用户关系; 另一方面, 多层循环 GCN 模块可提取并融合不同领域信息的共性与特性, 并通过高阶表示建模方式获取准确的用户偏好. 然而, 本文模型对重叠用户数据依赖性较强, 即当重叠用户交互信息匮乏时, 此时模型仅运用到多层循环 GCN 建模域内、域间关系, 难以发挥伪重叠判定机制的优势. 可见, 本文模型 PO-CDRec 在数据集较大、重叠用户交互数据较多的部分用户重叠场景下会有更好的表现.

3.4.2 消融实验分析 (RQ2)

为验证伪重叠判定机制和多层循环 GCN 这两个构件对跨域推荐模型性能的影响, 设计了 3 种变体模型, 分别是 Ours-1、Ours-2 和 Ours-3, 将它们与本文模型 PO-CDRec 进行对比. 其中, 变体模型 Ours-1 不使用伪重叠判定机制与多层循环 GCN, 通过 Louvain 算法分别对两域数据聚类, 并使用传统 GCN 学习节点特征, 主要验证两个构

件的作用; 变体模型 Ours-2 不采用伪重叠判定机制对跨域用户聚类, 而是对每个数据域使用 Louvain 算法聚类相似用户群体, 并通过多层循环 GCN 捕获用户偏好, 主要验证多层循环 GCN 构件获取用户偏好的有效性; 变体模型 Ours-3 不使用多层循环 GCN, 而仅通过固定图结构学习节点特征, 不考虑不同域间节点的关系, 主要阐明伪重叠判定机制构件对跨域推荐聚类效果的准确性. 各模型的构件描述情况如表 7 所示.

表 7 各模型的构件描述情况

变体模型	伪重叠判定机制	多层循环GCN
Ours-1	×	×
Ours-2	×	√
Ours-3	√	×
PO-CDRec	√	√

在 Amazon 数据集中选择 4 组数据域进行消融实验, 分别是: Music-Movie、Phone-Elec、Cloth-Sport、Game-Video, 考虑到变体模型的目的是验证模型构件对推荐性能的影响, 这里选择 MRR 、 $NDCG@10$ 、 $HR@10$ 作为评价指标. 实验结果如表 8 所示, 其中, 为便于比较, 利用下划线来突显变体模型中表现最佳的数据, 最后一列给出本文模型相对于某一最佳变体模型的性能对比情况.

表 8 各变体模型在不同数据集上的指标对比

数据集	评价指标	模型变体				性能对比情况 (%)
		Ours-1	Ours-2	Ours-3	PO-CDRec	
Music	MRR	4.30	5.95	<u>6.91</u>	7.13	↑3.09
	$NDCG@10$	4.60	6.59	<u>7.32</u>	7.91	↑7.46
	$HR@10$	9.23	13.30	<u>14.06</u>	14.98	↑6.14
Movie	MRR	4.33	5.57	<u>6.47</u>	6.64	↑2.56
	$NDCG@10$	4.80	6.18	<u>7.08</u>	7.41	↑4.45
	$HR@10$	10.07	<u>14.23</u>	14.21	14.89	↑4.43
Phone	MRR	6.70	6.31	<u>8.10</u>	8.08	↓2.48
	$NDCG@10$	7.88	7.22	<u>9.00</u>	9.32	↑3.43
	$HR@10$	14.25	<u>17.12</u>	14.15	17.90	↑4.36
Elec	MRR	8.22	9.54	<u>10.31</u>	11.28	↑8.60
	$NDCG@10$	9.44	10.96	<u>11.56</u>	11.63	↑0.60
	$HR@10$	16.58	18.50	<u>19.89</u>	20.43	↑2.64
Cloth	MRR	3.56	4.31	<u>5.72</u>	6.28	↑8.92
	$NDCG@10$	3.98	4.78	<u>5.77</u>	5.84	↑1.20
	$HR@10$	7.54	9.97	<u>11.26</u>	11.89	↑5.30
Sport	MRR	3.68	4.16	<u>4.97</u>	5.18	↑4.05
	$NDCG@10$	4.02	4.52	<u>5.67</u>	5.87	↑3.41
	$HR@10$	7.87	9.17	<u>11.36</u>	11.42	↑0.53
Game	MRR	2.99	3.72	<u>4.51</u>	4.87	↑7.40
	$NDCG@10$	2.78	3.98	<u>4.12</u>	4.92	↑16.26
	$HR@10$	5.60	7.64	<u>8.06</u>	9.17	↑12.10
Video	MRR	3.79	4.54	<u>4.69</u>	5.02	↑6.57
	$NDCG@10$	3.93	4.65	<u>5.28</u>	5.68	↑7.04
	$HR@10$	7.48	10.26	<u>11.32</u>	12.18	↑7.06

根据表 8, 各构件的实验分析如下: 伪重叠判定机制对模型性能影响较大, 传统的跨域推荐聚类方法在 MRR 、 $NDCG@10$ 、 $HR@10$ 这 3 个指标上的表现均不如伪重叠判定机制方法; 固定图结构 GCN 学习节点特征在这 3 个指标上的表现均不如多层循环 GCN 方法; 而去除伪重叠判定机制或多层循环 GCN 单个构件的模型与将二者结合的本文模型相比, 其效果明显更差. 下面给出更详细的分析.

3.4.2.1 通过多层循环 GCN 动态学习用户偏好对模型性能的影响较大

传统基于 GCN 的 CDR 算法在获取高阶节点嵌入前, 往往采用固定的图形式构建用户-项目节点关系, 构建过程将忽略不同域间节点关系的变化. 本文提出利用动态图结构分别学习域共享信息和域特有信息以获取高阶的跨域节点嵌入, 从而提高数据利用率, 缓解数据稀疏问题. 由表 8 可看出, 与变体模型 Ours-1 相比, 变体模型 Ours-2 在性能上提升较大. 比如, 在 MRR 指标上, 平均提升 15.34%, 最优时提升 27.73%. 由此可知, 基于 GCN 的 CDR 算法中增加循环结构并使图动态化, 可有效提高节点的鲁棒性和数据利用率, 从而获取用户最新偏好, 有利于提升模型推荐性能.

3.4.2.2 基于伪重叠判定机制聚合跨域用户对推荐性能的影响很大

跨域推荐中不同领域之间具有复杂关系, 而传统跨域推荐大多套用适用于单域推荐的聚类算法, 无法有效利用重叠信息学习域共享嵌入, 从而为后续特征学习提供更准确的聚类结果. 因此, 需针对部分用户重叠推荐场景对聚类算法进行改进. 对此, 所提模型 PO-CDRec 设计伪重叠判定机制, 整合聚类用户关系. 由表 8 可看出, 与变体模型 Ours-1 相比, 变体模型 Ours-3 在性能上提升明显. 比如, 在 HR 指标上, 平均提升 25.93%, 最优时提升 34.35%. 由此可知, 采用伪重叠判定机制能有效地利用重叠信息处理不同领域的复杂用户关系, 为后续特征学习提供更准确的聚类结果, 进而提高推荐准确性.

3.4.2.3 两构件相互配合使得模型推荐性能达到最优

与所有变体模型 Ours-1、Ours-2、Ours-3 相比, 本文模型 PO-CDRec 同时运用伪重叠判定机制与多层循环 GCN 这两个构件, 其评价指标在各数据域中普遍有所提升, MRR 、 $NDCG@10$ 、 $HR@10$ 分别平均提升 4.84%、5.48%、5.32%. 由此可知, 这两个构件相互配合可使 PO-CDRec 模型推荐效果达到最优.

由上述分析可知, 伪重叠判定机制以及多层循环 GCN 构件均对模型性能提升起到一定作用, 且二者的结合可更有效地增强模型在 MRR 、 $NDCG@10$ 、 $HR@10$ 指标上的效果.

3.4.3 参数敏感度分析 (RQ3)

这里对两个关键参数展开实验讨论, 即传播深度和重叠用户影响权重, 分析它们在不同数据集上对模型性能的影响.

3.4.3.1 传播深度

我们设置传播深度为参数 K , $K \in \{0, 1, 2, 3, 4\}$. 当 $K=0$ 时, 图卷积部分消失, 多层循环 GCN 模块退化为贝叶斯个性化排序 (Bayesian personalized ranking, BPR). 不同传播层的实验结果如图 4 所示.

由图 4 可知, 当 K 值为 0 时, 此时推荐指标均较小, 推荐性能较差. 当 K 由 0 变为 1 时, 模型性能提升明显, 表明 GCN 层可有效解决数据稀疏问题. 随着 K 的不断增大, 模型性能并未呈现不断增涨趋势, 而是在某个值之后下降. 在数据域 Music-Movie、Cloth-Sport、Game-Video 中 $K=3$ 时以及数据域 Phone-Elec 中 $K=2$ 时模型性能达到最佳. 原因在于, 对于稀疏数据集, 更深的图卷积可帮助聚合更多的邻居, 利于用户嵌入的学习, 从而提高冷启动用户的推荐准确率; 而对于密集数据集, 过高的传播层深度易导致 GNN 出现过度平滑现象.

3.4.3.2 重叠用户影响权重

为获取伪重叠用户社区, 使用 λ 作为重叠用户影响权重, 以衡量伪重叠判定机制所构建的用户社区对用户聚类准确性的影响, 其值在 $\{0.005, 0.01, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5\}$ 中选取. 将 MRR 、 $NDCG@10$ 作为评价指标, 验证 λ 的不同取值对推荐效果的影响, 如图 5 所示.

由图 5 可知, 随着 λ 值的变化, 不同数据域下的推荐指标 MRR 和 $NDCG@10$ 呈现先增后降的趋势. 第 3.4.2 节消融实验中的变体模型 Ours-2 可看作 λ 值为 0 时的情况, 此时推荐指标均为最小, 推荐性能较差. 在数据域 Music-Movie 中, 当重叠用户影响权重取值为 0.25 时, 推荐性能达到最佳; 在数据域 Cloth-Sport、Game-Video、Phone-Elec 中, λ 值为 0.3 时, 推荐性能达到最佳. 这表明, 通过调整 λ 值的大小, 可优化推荐性能.

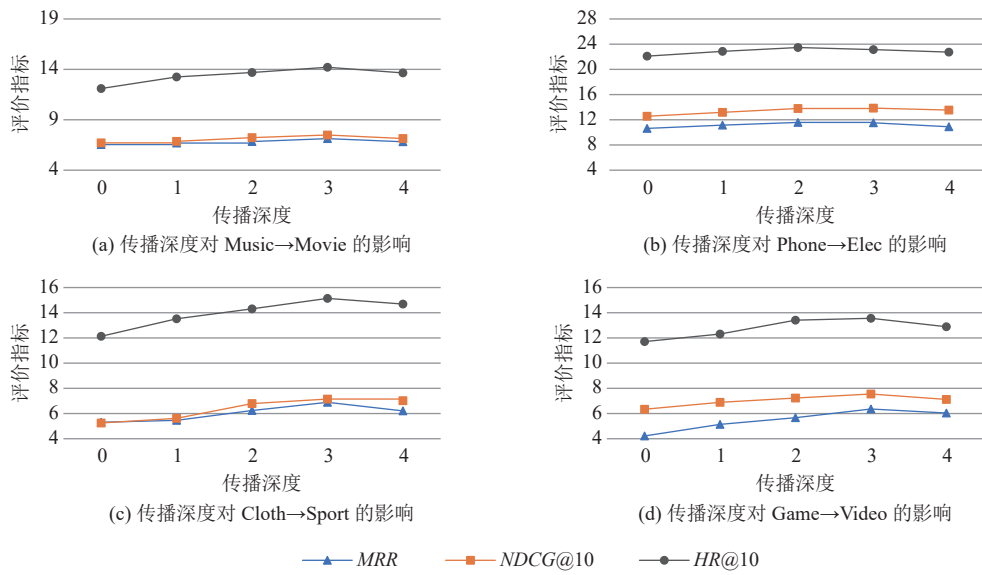


图4 传播深度的影响

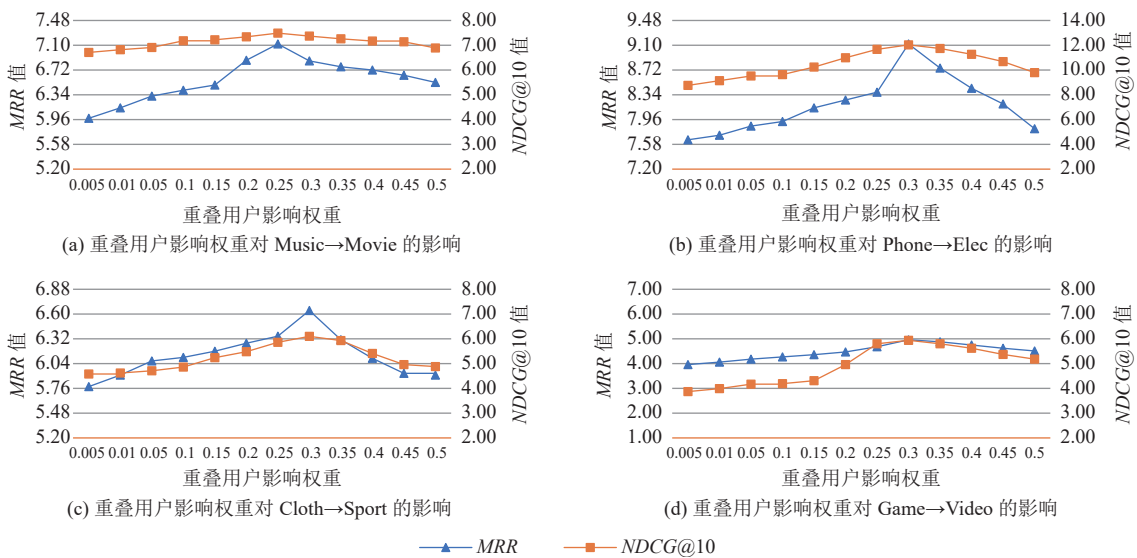


图5 重叠用户影响权重的影响

3.4.4 模型复杂度分析

为充分获取重叠用户信息, 所提模型 PO-CDRec 在社区聚类算法 Louvain 的基础上设计伪重叠判定机制, 据此挖掘用户的信任关系以及相似用户社区, 再结合提出的包含嵌入学习模块和图学习模块的多层循环 GCN, 学习动态的域共享特征、域特有特征以及动态图结构, 并通过两模块的循环增强, 获取最新用户偏好, 进而对用户推荐项目. 与现有相关模型相比, PO-CDRec 的推荐过程稍长, 复杂度相对有所增加, 但推荐效果得到较大提升 (见表 3-表 6). 下面对本文模型及部分具有代表性的对比模型就其时间复杂度进行分析.

1) 就本文模型 PO-CDRec 而言, 其时间复杂度主要表现在伪重叠判定机制用户聚类与多层循环 GCN 偏好学习两个部分. 伪重叠判定机制用户聚类部分主要利用社区聚类算法. 在局部优化阶段, 每个节点均需计算其加入相

邻节点所在社区后的模块度变化,并选择最优的社区进行合并,时间复杂度为 $O(NM)$,其中 N 是用户数, M 是邻居节点数. 经过局部优化后,每个社区被视为一个新的超节点,形成新的简化网络,新网络的节点数减少为原网络的社区数,这通常显著小于 N . 因网络规模在每次合并后缩小,故社区合并阶段的迭代次数在最坏情况下为 $O(\log N)$. 可知,伪重叠判定机制用户聚类部分的时间复杂度大致为 $O(NM \log N)$. 对于多层循环 GCN 部分,其运算代价主要体现在图卷积操作中. 单层图卷积操作的时间复杂度为 $O(FEd)$,其中 F 为每层的节点数,每个节点的平均邻居数为 E , d 是每个节点的特征维度,对于每一层的图卷积操作需遍历每个节点及其邻居,并进行相应运算. 可知,整个多层循环 GCN 部分的时间复杂度为 $O(KFEd)$, K 为嵌入传播深度. 因此,本文模型的时间复杂度大致为 $O(NM \log N + KFEd)$.

2) 就对比模型 HCCDR(相对较新的改进聚类算法的跨域推荐模型)的时间复杂度来说,其时间主要消耗在软聚类分配节点上,时间复杂度大致为 $O(LNId)$, L 为平均聚类簇数, I 为训练轮数. 与模型 HCCDR 相比,本文模型 PO-CDRec 的复杂度虽有所增加,但模型 HCCDR 仅侧重用户聚类,导致模型适应性不足. 且就目前大多数 GPU/CPU 来说,本文模型的计算任务并不算大. 特别地,从推荐效果来看,所提模型 PO-CDRec 比 HCCDR 的性能有所提升.

3) 就对比模型 FPPDM (在仅改进聚类算法的跨域推荐模型中表现最优)的时间复杂度来说,其时间主要消耗在单域建模、全局聚合以及紧致性联合聚类上,时间复杂度大致为 $O(IEd + cgN + LNId)$,其中 c 表示参与联邦学习的客户端数量, g 表示全局聚合轮数. 相较于本文模型而言, FPPDM 模型使用联邦学习涉及多个本地域之间的模型参数传输和聚合,更侧重于隐私保护,需消耗更多的时间和计算资源,导致其时间复杂度较高.

4) 就对比模型 I^2 RCDR (相对较新的基于 GCN 的跨域推荐模型)的时间复杂度来说,其时间主要消耗在图卷积网络上,时间复杂度大致为 $O(IEFd)$. 模型 I^2 RCDR 针对重叠用户的跨域场景,更侧重于用户关系的挖掘,忽视了非重叠用户场景,导致用户相似度迁移不准确. 与模型 I^2 RCDR 相比,本文模型 PO-CDRec 的复杂度虽有所增加,但换来了较大的性能提升.

5) 就对比模型 GLOU (在对比模型中平均性能最佳)的时间复杂度来说,其时间主要由图卷积网络的计算决定,时间复杂度大致为 $O(NMI + LFE d)$. 该模型根据偏好对用户分组并利用 GCN 挖掘不同用户组图中的节点高阶关系. 因在通常情况下,平均聚类簇数 L 远大于嵌入传播深度 K ,故本文模型 PO-CDRec 的复杂度通常会低于对比模型 GLOU. 另外,由对比实验结果可知,本文模型 PO-CDRec 的性能也较之更优.

综上所述,各模型的时间复杂度大小关系为: $GLOU > FPPDM > PO-CDRec > I^2RCDR > HCCDR$. 尽管从时间复杂度上来看,本文模型 PO-CDRec 要劣于部分对比模型,但就推荐性能而言,本文模型 PO-CDRec 优于 HCCDR 和 I^2 RCDR. 相较于 HCCDR,所提模型在 MRR 、 $NDCG@N$ 和 $HR@N$ 指标上分别最少提升 5.87%, 5.68%, 5.37%. 相较于 I^2 RCDR,所提模型在 MRR 、 $NDCG@N$ 和 $HR@N$ 指标上分别最少提升 10.09%, 13.42%, 11.09%. 相较于 GLOU 和 FPPDM 而言,所提模型在推荐性能和时间复杂度方面均占优. 其中,与对比模型中平均性能最佳的 GLOU 相比,所提模型 PO-CDRec 在所有数据域上的指标 MRR 、 $NDCG@N$ 和 $HR@N$ 分别最少提升 2.65%, 1.26%, 0.77%. 可见,本文模型的设计思路是合理的、有意义的.

4 总结与下一步工作

本文提出一种利用伪重叠判定机制的多层循环 GCN 跨域推荐模型 PO-CDRec. 该模型运用伪重叠判定机制挖掘用户的信任关系以及用户社区,获取用户相似性,从而提高聚类算法在跨域推荐中的准确率;同时采用多层循环 GCN 学习节点嵌入,充分挖掘节点高阶关系以缓解数据稀疏现象;通过门控神经单元生成用户与项目的组合特征,利用 MLP 获取源域与目标域最终预测得分概率,达到更佳的推荐效果.

1) 为提高模型推荐准确性,设计伪重叠判定机制. 为充分获取重叠用户信息,并使聚类算法更适应于跨域推荐中,所提模型设计伪重叠用户判断机制联合源域和目标域的用户进行聚类,生成用户社交网络,得到信任关系和用

户相似度, 从而提高聚类算法适应能力及推荐准确性.

2) 为缓解数据稀疏问题, 运用多层循环 GCN. 以动态图结构代替传统图卷积中的固定图学习, 同时从域内/域间角度出发, 充分利用不同域数据信息聚合节点嵌入, 增强模型的高阶连通性, 缓解数据稀疏现象.

3) 为验证所提模型效果, 展开综合实验分析. 无论是与经典模型还是最新模型对比, 所提模型 PO-CDRec 在 4 种跨域推荐场景中, 其 3 个常用推荐指标均有明显提升. 同时, 通过消融实验, 验证了模型各组件的必要性. 最后, 开展参数敏感度实验, 优化了模型的关键参数.

所提模型 PO-CDRec 虽表现不错, 但依然存在一些待完善之处, 其中伪重叠判定机制未考虑对不同重叠用户占比的数据集作相应改进. 在下一步工作中, 我们将考虑设计动态伪重叠判断机制, 以区分重叠用户和伪重叠用户, 进一步提升模型性能. 另外, 模型的可解释性也一直备受关注, 我们还将引入特征重要性 (如用户/项目特征权重、贡献度等) 以试图增强模型透明性与推荐公平性.

References:

- [1] He XN, Liao LZ, Zhang HW, Nie LQ, Hu X, Chua TS. Neural collaborative filtering. In: Proc. of the 26th Int'l Conf. on World Wide Web. Perth: ACM, 2017. 173–182. [doi: [10.1145/3038912.3052569](https://doi.org/10.1145/3038912.3052569)]
- [2] Zhu F, Chen CC, Wang Y, Liu GF, Zheng XL. DTCDR: A framework for dual-target cross-domain recommendation. In: Proc. of the 28th ACM Int'l Conf. on Information and Knowledge Management. Beijing: ACM, 2019. 1533–1542. [doi: [10.1145/3357384.3357992](https://doi.org/10.1145/3357384.3357992)]
- [3] Zhu F, Wang Y, Zhou J, Chen CC, Li LF, Liu GF. A unified framework for cross-domain and cross-system recommendations. IEEE Trans. on Knowledge and Data Engineering, 2023, 35(2): 1171–1184. [doi: [10.1109/TKDE.2021.3104873](https://doi.org/10.1109/TKDE.2021.3104873)]
- [4] Wang LE, Li DC, Li XX. Deep recommendation model with cross-domain association and privacy protection. Ruan Jian Xue Bao/Journal of Software, 2023, 34(7): 3365–3384 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/6533.htm> [doi: [10.13328/j.cnki.jos.006533](https://doi.org/10.13328/j.cnki.jos.006533)]
- [5] Perera D, Zimmermann R. Towards comprehensive recommender systems: Time-aware unified recommendations based on listwise ranking of implicit cross-network data. In: Proc. of the 34th AAAI Conf. on Artificial Intelligence. New York: AAAI, 2020. 189–197. [doi: [10.1609/aaai.v34i01.5350](https://doi.org/10.1609/aaai.v34i01.5350)]
- [6] Zhang JY, Wa CX, Li C, Zhao ZY. Towards lightweight cross-domain sequential recommendation via tri-branches graph external attention network. Journal of Computer Research and Development, 2024, 61(8): 1930–1944 (in Chinese with English abstract). [doi: [10.7544/issn1000-1239.202440197](https://doi.org/10.7544/issn1000-1239.202440197)]
- [7] Dong X, Yu L, Wu ZH, Sun YX, Yuan LF, Zhang FX. A hybrid collaborative filtering model with deep structure for recommender systems. In: Proc. of the 31st AAAI Conf. on Artificial Intelligence. San Francisco: AAAI, 2017. 1309–1315. [doi: [10.1609/aaai.v31i1.10747](https://doi.org/10.1609/aaai.v31i1.10747)]
- [8] Zhou KX, Huang X, Li YN, Zha DC, Chen R, Hu X. Towards deeper graph neural networks with differentiable group normalization. In: Proc. of the 34th Int'l Conf. on Neural Information Processing Systems. Vancouver: ACM, 2020. 413.
- [9] Huang L, Huang ZW, Huang ZY, Guan CR, Gao YF, Wang CD. Graph convolutional broad cross-domain recommender system. Journal of Computer Research and Development, 2024, 61(7): 1713–1729. (in Chinese with English abstract). [doi: [10.7544/issn1000-1239.202330617](https://doi.org/10.7544/issn1000-1239.202330617)]
- [10] Gao C, Zheng Y, Li N, Li YF, Qin YR, Piao JH, Quan YH, Chang JX, Jin DP, He XN, Li Y. A survey of graph neural networks for recommender systems: challenges, methods, and directions. ACM Trans. on Recommender Systems, 2023, 1(1): 3. [doi: [10.1145/3568022](https://doi.org/10.1145/3568022)]
- [11] Jin H, Hu ZQ. The non-negative matrix factorization based algorithm for community detection in sparse networks. Acta Electronica Sinica, 2023, 51(10): 2950–2959 (in Chinese with English abstract). [doi: [10.12263/DZXB.20210950](https://doi.org/10.12263/DZXB.20210950)]
- [12] Afoudi Y, Lazaar M, Al Achhab M. Hybrid recommendation system combined content-based filtering and collaborative prediction using artificial neural network. Simulation Modelling Practice and Theory, 2021, 113: 102375. [doi: [10.1016/j.simpat.2021.102375](https://doi.org/10.1016/j.simpat.2021.102375)]
- [13] Siet S, Peng S, Ilkhomejion S, Kang MS, Park DS. Enhancing sequence movie recommendation system using deep learning and kmeans. Applied Sciences, 2024, 14(6): 2505. [doi: [10.3390/app14062505](https://doi.org/10.3390/app14062505)]
- [14] Chang JX, Gao C, Zheng Y, Hui YQ, Niu YN, Song Y, Jin DP, Li Y. Sequential recommendation with graph neural networks. In: Proc. of the 44th Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval. New York: ACM, 2021. 378–387. [doi: [10.1145/3404835.3462968](https://doi.org/10.1145/3404835.3462968)]

- [15] Kang SK, Hwang J, Lee D, Yu H. Semi-supervised learning for cross-domain recommendation to cold-start users. In: Proc. of the 28th ACM Int'l Conf. on Information and Knowledge Management. Beijing: ACM, 2019. 1563–1572. [doi: 10.1145/3357384.3357914]
- [16] Xu J, Wang X, Zhang HM, Lv P. Heterogeneous and clustering-enhanced personalized preference transfer for cross-domain recommendation. *Information Fusion*, 2023, 99: 101892. [doi: 10.1016/j.inffus.2023.101892]
- [17] Liu WM, Chen CC, Liao XT, Hu ML, Yin JW, Tan YC, Zheng LF. Federated probabilistic preference distribution modelling with compactness co-clustering for privacy-preserving multi-domain recommendation. In: Proc. of the 32nd Int'l Joint Conf. on Artificial Intelligence. Macao: IJCAI, 2023. 2206–2214. [doi: 10.24963/ijcai.2023/245]
- [18] Wang X, He XN, Nie LQ, Chua TS. Item silk road: recommending items from information domains to social users. In: Proc. of the 40th Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval. Shinjuku: ACM, 2017. 185–194. [doi: 10.1145/3077136.3080771]
- [19] Cui Q, Wei T, Zhang YF, Zhang Q. HeroGRAPH: a heterogeneous graph framework for multi-target cross-domain recommendation. In: Proc. of the 3rd Workshop on Online Recommender Systems and User Modeling Co-located with the 14th ACM Conf. on Recommender Systems. New York: CEUR, 2020.
- [20] Xu K, Xie YZ, Chen L, Zheng ZB. Expanding relationship for cross domain recommendation. In: Proc. of the 30th ACM Int'l Conf. on Information & Knowledge Management. New York: ACM, 2021. 2251–2260. [doi: 10.1145/3459637.3482429]
- [21] Li J, Peng ZH, Wang SZ, Xu XK, Yu PS, Hao ZY. Heterogeneous graph embedding for cross-domain recommendation through adversarial learning. In: Proc. of the 25th Int'l Conf. on Database Systems for Advanced Applications. Jeju: Springer, 2020. 507–522. [doi: 10.1007/978-3-030-59419-0_31]
- [22] Cao JX, Sheng JW, Cong X, Liu TW, Wang B. Cross-domain recommendation to cold-start users via variational information bottleneck. In: Proc. of the 38th Int'l Conf. on Data Engineering. Kuala Lumpur: IEEE, 2022. 2209–2223. [doi: 10.1109/ICDE53745.2022.00211]
- [23] Mutiara, Mutiara AB, Wirawan S, Yusnitasari T, Anggraini D. Expanding louvain algorithm for clustering relationship formation. *Int'l Journal of Advanced Computer Science and Applications*, 2023, 14(1): 701–708. [doi: 10.14569/IJACSA.2023.0140177]
- [24] Liu Q, Cheng Y. Research on multi-granularity neural network pruning method with regularization mechanism. *Acta Electronica Sinica*, 2023, 51(8): 2202–2212 (in Chinese with English abstract). [doi: 10.12263/DZXB.20210844]
- [25] Zhou JK, Wang N, Cui L. EasiLTOM: signal activity interval recognition based on local dynamic threshold. *Journal of Computer Research and Development*, 2022, 59(4): 826–833 (in Chinese with English abstract). [doi: 10.7544/issn1000-1239.20200662]
- [26] Hsieh CK, Yang LQ, Cui Y, Lin TY, Belongie S, Estrin D. Collaborative metric learning. In: Proc. of the 26th Int'l Conf. on World Wide Web. Perth: ACM, 2017. 193–201. [doi: 10.1145/3038912.3052639]
- [27] Cui ZH, Xu XH, Xue F, Cai XJ, Cao Y, Zhang WS, Chen JJ. Personalized recommendation system based on collaborative filtering for IoT scenarios. *IEEE Trans. on Services Computing*, 2020, 13(4): 685–695. [doi: 10.1109/TSC.2020.2964552]
- [28] Wang X, He XN, Wang M, Feng FL, Chua TS. Neural graph collaborative filtering. In: Proc. of the 42nd Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval. Paris: ACM, 2019. 165–174. [doi: 10.1145/3331184.3331267]
- [29] Man T, Shen HW, Jin XL, Cheng XQ. Cross-domain recommendation: an embedding and mapping approach. In: Proc. of the 26th Int'l Joint Conf. on Artificial Intelligence. Melbourne: IJCAI, 2017. 2464–2470. [doi: 10.24963/ijcai.2017/343]
- [30] Sheng XR, Zhao LQ, Zhou GR, Ding XY, Dai BD, Luo Q, Yang SR, Lv JS, Zhang C, Deng HB, Zhu XQ. One model to serve all: Star topology adaptive recommender for multi-domain CTR prediction. In: Proc. of the 30th ACM Int'l Conf. on Information & Knowledge Management. New York: ACM, 2021. 4104–4113. [doi: 10.1145/3459637.3481941]
- [31] Xu JJ, Song JY, Sang Y, Yin LH. CDAML: A cluster-based domain adaptive meta-learning model for cross domain recommendation. *World Wide Web*, 2023, 26(3): 989–1003. [doi: 10.1007/s11280-022-01068-5]
- [32] Zhao C, Li CL, Fu C. Cross-domain recommendation via preference propagation graphnet. In: Proc. of the 28th ACM Int'l Conf. on Information and Knowledge Management. Beijing: ACM, 2019. 2165–2168. [doi: 10.1145/3357384.3358166]
- [33] Wang K, Zhu YM, Liu HB, Zang TZ, Wang CY, Liu K. Inter- and intra-domain relation-aware heterogeneous graph convolutional networks for cross-domain recommendation. In: Proc. of the 27th Int'l Conf. on Database Systems for Advanced Applications. Berlin: Springer, 2022. 53–68. [doi: 10.1007/978-3-031-00126-0_4]
- [34] Liu M, Li JJ, Guo ZQ, Li GH. Extracting latently overlapping users by graph neural network for non-overlapping cross-domain recommendation. *Knowledge-based Systems*, 2024, 290: 111508. [doi: 10.1016/j.knosys.2024.111508]

附中文参考文献:

- [4] 王利娥, 李东城, 李先贤. 基于跨域关联与隐私保护的深度推荐模型. *软件学报*, 2023, 34(7): 3365–3384. <http://www.jos.org.cn/1000->

[9825/6533.htm](#) [doi: [10.13328/j.cnki.jos.006533](#)]

- [6] 张劲羽, 马晨曦, 李超, 赵中英. 基于三支图外部注意力网络的轻量化跨域序列推荐. 计算机研究与发展, 2024, 61(8): 1930–1944. [doi: [10.7544/issn1000-1239.202440197](#)]
- [9] 黄玲, 黄镇伟, 黄梓源, 关灿荣, 高月芳, 王昌栋. 图卷积宽度跨域推荐系统. 计算机研究与发展, 2024, 61(7): 1713–1729. [doi: [10.7544/issn1000-1239.202330617](#)]
- [11] 金红, 胡智群. 基于非负矩阵分解的稀疏网络社区发现算法. 电子学报, 2023, 51(10): 2950–2959. [doi: [10.12263/DZXB.20210950](#)]
- [24] 刘奇, 陈莹. 正则化机制下多粒度神经网络剪枝方法研究. 电子学报, 2023, 51(8): 2202–2212. [doi: [10.12263/DZXB.20210844](#)]
- [25] 周钧锴, 王念, 崔莉. EasiLTOM: 一种基于局部动态阈值的信号活动区间识别方法. 计算机研究与发展, 2022, 59(4): 826–833. [doi: [10.7544/issn1000-1239.20200662](#)]



钱忠胜(1977—), 男, 博士, 教授, 博士生导师, CCF 高级会员, 主要研究领域为软件工程, 机器学习, 智能化软件.



范赋宇(2002—), 男, 硕士生, CCF 学生会员, 主要研究领域为软件工程, 机器学习.



王亚惠(1999—), 女, 硕士生, 主要研究领域为软件工程, 机器学习.



付庭峰(1999—), 男, 硕士生, 主要研究领域为软件工程, 机器学习.



俞情媛(1997—), 女, 博士生, 主要研究领域为软件工程, 机器学习.