山西大学学报(自然科学版)49(4):557—570,2026
Journal of Shanxi University(Nat. Sci. Ed.)

DOI:10.13451/j.sxu.ns.2025111

层次拓扑增强的多模态推荐方法

许钊菁[1],王海荣[1*],易之航[1],王静[2,3,4]

(1.北方民族大学 计算机科学与工程学院,宁夏 银川 750021;
2.中国气象局旱区特色农业气象灾害监测预警与风险管理重点实验室,宁夏 银川 750002;
3.宁夏气象防灾减灾重点实验室,宁夏 银川 750002;
4.宁夏回族自治区气象科学研究所,宁夏 银川 750002)

摘 要:针对推荐系统存在的多模态信息利用不充分和交互数据噪声问题,本文从交互去噪的角度出发,提出层次拓扑增强的多模态推荐方法。该方法通过层次锚点拓扑结构嵌入,按节点活跃度分层采样锚点并建模全局关联性,解决过度平滑和长距离依赖捕捉问题,基于自监督学习设计可学习的生成器来生成用户的多模态偏好与去噪交互图,进行用户偏好和交互图的双向协同更新,设计"图结构-模态"一致性约束有效抑制交互噪声。在三个真实数据集上实验,将其结果与13个基线模型对比,本文方法在Recall@20和NDCG@20指标上分别平均提升了2.37%和3.27%。

关键词:自监督学习;图神经网络;交互去噪

中图分类号:TP39 **文献标志码**:A **文章编号**:0253-2395(2026)04-0557-14

Hierarchical Topology-enhanced Multi-modal Recommendation Method

Xu Zhaojing[1], Wang Hairong[1*], Yi Zhihang[1], Wang Jing[2,3,4]

(1. *School of Computer Science and Engineering, North Minzu University, Yinchuan* 750021*, China;*
2. *Key Laboratory of Characteristic Agrometeorological Disaster Monitoring, Early Warning and Risk Management in Arid Regions China Meteorological Administration, Yinchuan* 750002*, China;*
3. *Ningxia Key Lab of Meteorological Disaster Prevention and Reduction, Yinchuan* 750002*, China;*
4. *Institute of Meteorological Sciences in Ningxia Hui Autonomous Region, Yinchuan* 750002*, China*)

Abstract: Recommendation systems often suffer from insufficient utilization of multi-modal information and noisy interaction data. To address these issues, this paper proposes a hierarchical topology-enhanced multi-modal recommendation method from an interaction denoising perspective. The method employs hierarchical anchor topology embedding, which performs stratified anchor sampling based on node activity levels to model global relationships, thereby mitigating over-smoothing and capturing long-range dependencies. Based on self-supervised learning, a learnable generator is designed to produce user multi-modal preferences and denoised interaction graphs, enabling bidirectional collaborative updating between user preferences and interaction graphs. A "graph-structure-modality" consistency constraint is incorporated to effectively suppress interaction noise. Experiments conducted on three real-world datasets demonstrate that the proposed method achieves average improvements of 2.37% and 3.27% on Recall@20 and NDCG@20

收稿日期:2025-09-12;**修回日期**:2025-09-28

基金项目:北方民族大学研究生教育质量提升项目(YJZT202424);中国气象局旱区特色农业气象灾害监测预警与风险管理重点实验室宁夏气象防灾减灾重点实验室开放研究项目(CAMF-202502);北方民族大学研究生创新项目(CYX25221)

作者简介:许钊菁(2002—),女,宁夏银川人,硕士研究生,研究方向为多模态推荐算法、知识图谱。E-mail:20247555@stu.nmu.edu.cn

* **通信作者**:王海荣(Wang Hairong),E-mail: wanghr@nun.edu.cn

引文格式:许钊菁,王海荣,易之航,等.层次拓扑增强的多模态推荐方法[J].山西大学学报(自然科学版),2026,49(4):557-570. DOI:10.13451/j.sxu.ns.2025111.

metrics, respectively, compared with 13 baseline models.

Key words: self-supervised learning; graph neural networks; interaction denoising

0 引言

推荐系统作为信息过载问题的核心解决方案,其研究始终围绕精准捕捉用户潜在偏好与复杂交互关系展开[1]。从早期基于协同过滤的浅层模型到融合多模态信息与图结构的深度框架,研究范式经历了多次重要革新。

以协同过滤[2]为代表的传统推荐系统,通过用户-项目交互矩阵的低秩近似实现推荐[3]。这类方法假设用户与项目特征可通过线性组合建模,如基于矩阵分解的贝叶斯个性化排序(Matrix Factorization with Bayesian Personalized Ranking,MF-BPR)模型[4]借助贝叶斯个性化排序优化矩阵分解过程,在数据稀疏场景中展现出一定有效性。然而,此类方法仅能利用低阶交互信息,无法对用户-项目交互网络的高阶连通性进行建模,难以刻画用户的复杂行为模式。

随着图神经网络(Graph Neural Network,GNN)[5]在图结构建模中的兴起,推荐系统迎来了突破性进展,如神经图协同过滤(Neural Graph Collaborative Filtering,NGCF)[6]等通过消息传递机制聚合邻域特征的推荐模型。这类方法能够有效地捕获交互图中的高阶协同信号[7],但其过度平滑问题[8]使节点特征在深层传播中逐渐趋于同质化,同时误点击、虚假交互则会误导模型学习,降低推荐精度。

针对该问题,自监督学习(Self-Supervised Learning,SSL)[9]的引入成为新方向。自监督学习通过数据增强和构建对比学习样本或对比学习视图,迫使相似用户或项目的嵌入在特征空间中彼此靠近,从而增强表征的鲁棒性。但这类方法仍受限于静态图结构和拓扑特征的局限性,固定交互图难以动态适应噪声干扰[10]。对于现有方法普遍存在视图生成与语义脱节、单模态信息瓶颈等问题,引入多模态数据作为协同信息的重要补充也被认为是可行的解决方案。

现实世界中的项目通常蕴含丰富的多模态信息(如视觉图像、文本描述、音频)。这些信息为理解项目语义和用户兴趣提供了宝贵的补充线索[11]。有效融合多模态信息,不仅能缓解数据稀疏性问题,更能揭示用户深层次的偏好,提升推荐的准确性和可解释性。但尽管多模态推荐系统借助文本、图像等辅助信息的整合,让推荐的语义丰富度得到了显著提升。在挖掘多模态数据中的有效信息时,怎样融合不同模态下的特征表示并规避噪声干扰,仍是现有研究中存在的明显局限问题。此外,当前的相关方法还存在模态信息利用不够充分、图结构与模态交互关系割裂的问题,制约了模型性能与推广能力。

为了解决上述问题,本文提出了一种层次拓扑增强的多模态推荐方法(Hierarchical Topology-Enhanced Multi-Modal Recommendation Method,HEMMR)。该方法通过层次锚点拓扑机制生成全局感知的节点表示以增强节点特征,改善过度平滑和长距离依赖捕捉困难问题。利用自编码器和多头注意力机制设计可学习的生成器,进行增强交互图和用户多模态偏好特征的生成。通过在两者间构造双向反馈通道,设计“图结构-模态”一致性约束和对比学习任务,能够有效抑制交互噪声。

1 相关工作

1.1 图神经网络在推荐系统中的应用

随着图神经网络的发展,推荐系统逐渐借助图结构来建模用户与项目的高阶关系,以缓解数据稀疏与冷启动等问题。轻量级图卷积网络(Light Graph Convolution Network,LightGCN)[12]通过简化的图卷积去除非线性变换,仅保留邻居特征聚合,大幅提升了协同信号的建模效率与准确性。

在此基础上,研究者逐渐将图神经网络扩展至多模态推荐场景,以更好地利用来自视觉、文本、语音等模态的辅助信息。多模态图卷积网络(Multimodal Graph Convolution Network,MGCN)[13]模型通过构建各模态下的用户-项目二部图,模拟用户对不同模态的关注,引导

信息传播。面向多模态推荐的隐结构挖掘方案(Latent Structure Mining Scheme for Multimodal Recommendation,LATREC)[14]通过模态感知结构学习层替代传统交互图,构建项目-项目关系图,从而捕捉多模态信息间的潜在关联。与此同时,注意力机制也被引入多模态推荐系统,可以在邻居信息聚合时分配权重,更加精细地刻画不同交互重要性。轻量级图 Transformer 模型(Light Graph Transformer Model,LightGT)[15]则采用轻量化图 Transformer,将注意力机制引入用户-项目建模中,更好地捕捉全局依赖。对于多模态信息在传播中的噪声问题,多视图图卷积网络(Multi-View Graph Convolutional Network,MVGCN)[16]通过多视图卷积架构,在各模态子图上执行独立的轻量图卷积以捕捉模态专属交互模式,并基于用户偏好调整模态特征的融合来缓解噪声污染和偏好不完整的问题。

然而,上述方法往往依赖较深的传播层数来覆盖高阶关系,容易导致过度平滑问题。同时,在多模态场景下,噪声信息在传播过程中可能被放大,削弱模型的鲁棒性。基于现有推荐方法在捕捉高阶关系的工作,本文设计层次锚点拓扑增强机制建模全局关联性,帮助去噪交互图生成器捕获全局结构信息,缓解过度平滑问题,使节点表征保持多样性与区分性。

1.2 自监督学习在推荐系统中的应用

自监督学习近年来在推荐系统中被广泛应用,其核心思路是通过数据的内在结构生成增强视图,并借助特定优化任务缓解数据稀疏性[17-18]。在单模态场景下,主要基于交互图展开。自监督图学习(Self-supervised Graph Learning,SGL)[19]创新性地采用节点丢弃、边丢弃和随机游走来构造对比视图,构造对比视图并引入自监督对比损失,从而提升嵌入表示的鲁棒性与区分性。与之不同,简易图对比学习方法(Simple Graph Contrastive Learning,SimGCL)[20]不再依赖传统的图结构增强策略,而是在嵌入空间中注入均匀噪声来构建对比视图,以增强特征鲁棒性。图 Transformer(Graph Transformer,GFormer)[21]则借鉴变分自编码器的思想,通过掩码和重建部分用户-项目交互信息,从而学习到更鲁棒的用户和物品表示。

在多模态推荐场景下,自监督学习方法聚焦于提升多模态特征的鲁棒性,核心思路是缓解模态内部特征噪声和模态融合过程中可能引入的信息偏差。多模态图对比学习(Multi-Modal Graph Contrastive Learning,MMGCL)[22]将自监督学习应用于微视频推荐,提出模态边缘丢弃、模态遮蔽和扰乱负样本模态等数据增强技术,确保各模态在推荐过程中充分发挥作用。自举多模态模型(Bootstrapped Multi-modal Model,BM3)[23]通过引入潜在嵌入随机丢弃结合自监督对比损失,降低对负采样及额外图增强的依赖与开销。自监督学习引导的多媒体推荐(Self-supervised Learning-guided Multimedia Recommendation,SLMRec)[24]则设计了基于多模态特征噪声扰动与多模态模式挖掘的自监督任务,并与推荐目标联合优化,以增强多模态推荐的鲁棒性与表达能力。多模态自监督学习(Multi-Modal Self-Supervised Learning,MMSSL)[25]通过模态感知对抗扰动与跨模态对比学习,有效捕捉不同模态下用户偏好的内在关联性。面向多模态推荐的局部与全局图学习(Local and Global Graph Learning for Multimodal Recommendation,LGMRec)[26]是一种基于局部和全局超图学习的多模态推荐系统,旨在同时建模用户的局部和全局兴趣,以提高推荐的准确性和鲁棒性。主图学习(Principal Graph Learning,PGL)[27]通过全局感知和局部感知子图从用户-项目交互图中提取主要子图,然后在主要子图上采用消息传递,以实现更精准的个性化推荐。

这些方法有效提升了多模态特征的鲁棒性,但其关注点主要局限于建模用户对模态信息的偏好和模态内部特征中的噪声控制,而对交互数据中固有的噪声考虑不足。基于上述自监督学习在多模态特征去噪和建模多模态偏好方面的应用,本文采用自编码器和多头注意力机制结合对比学习构建可学习的生成器,并探索多模态偏好能否提升多模态推荐模型在交互噪声下的鲁棒性。

2 方法模型

2.1 模型描述

层次拓扑增强的多模态推荐方法通过层次锚点拓扑结构增强用户、项目节点特征,设计可学习的生成器,用于联合建模用户的多模态偏好与去噪交互关系,利用对比学习任务进行自监督协同训练,构建“图结构-模态”一致性约束来抑制交互噪声。整体框架如图1所示。HEMMR主要由3个部分构成:

1)层次锚点拓扑嵌入。依据交互频度选择不同层次的锚点,分别增强用户、项目的节点特征并建模全局关联性;

2)多模态偏好辅助去噪交互图生成。通过多头注意力机制观测用户的潜在交互兴趣,为交互图中的边生成交互权重,该权重将用于构建去噪交互图生成器,同时通过多模态偏好协同进一步辅助去噪交互图的生成;

3)去噪交互图引导多模态偏好生成。借助生成的去噪交互图指导多模态偏好生成器,采用注意力机制生成用户多模态偏好并反向影响去噪交互图的生成,设计“图结构-模态”一致性约束,抑制交互噪声。

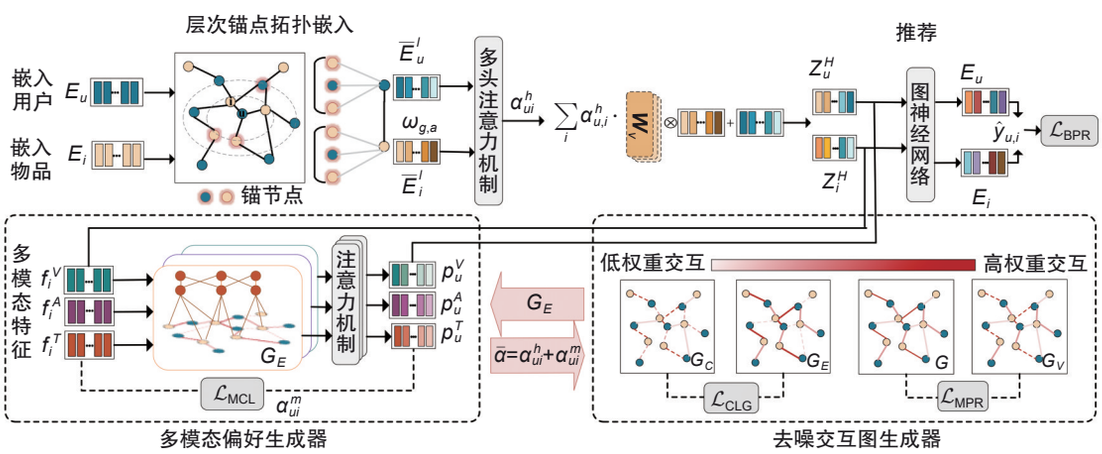

图1 HEMMR框架

注:HEMMR由层次锚点拓扑嵌入、多模态辅助去噪交互图生成、去噪图引导模态偏好生成三部分组成,实现双向协同更新。

Fig. 1 Framework of HEMMR

Note: HEMMR consists of three parts: hierarchical anchor topology embedding, multi-modal assisted denoising interaction graph generation, and denoised graph-guided modal preference generation, enabling bidirectional collaborative updating.

2.2 预备知识和符号定义

多模态推荐系统通过学习用户与项目的特征表示,预测用户对未交互项目的交互概率。为便于文中形式化描述,在表1列举文中所用到的主要符号及表示的含义。其中,多模态推荐任务各部分的相关定义如下。

定义1 多模态推荐。输入用户-项目交互图G、用户嵌入$\boldsymbol{E}_u$、项目嵌入$\boldsymbol{E}_i$和多模态特征$\boldsymbol{F}^m$,预测用户u对项目i的交互概率$\hat{y}_{ui}$。

定义2 交互图。对于用户集合U和项目集合I存在交互矩阵$\boldsymbol{Y}\in\mathbb{R}^{N_u\times N_i}$。若用户$u$与项目$i$交互,则$y_{ui}=1$,否则$y_{ui}=0$。用户与项目的交互形成图$G=(V,E)$,其中节点集$V=U\cup I$,边集$E=\{y_{ui}\mid y_{ui}=1\}$表示节点的交互。

定义3 多模态特征。项目的多模态特征由$\boldsymbol{F}^m=\{\boldsymbol{f}_1^m,\boldsymbol{f}_2^m,\cdots,\boldsymbol{f}_i^m,\cdots,\boldsymbol{f}_{N_i}^m\}$表示,$\boldsymbol{f}_i^m$表示经特征提取后项目$i$的第$m$种模态特征。

表1 主要符号说明

Table 1 Description of main symbols

符号	含义
$U=\{u_1,u_2,u_3,\cdots,u_{N_u}\}$	用户集合
$I=\{i_1,i_2,i_3,\cdots,i_{N_i}\}$	项目集合
$\boldsymbol{Y}$	交互矩阵
$G=(V,E)$	交互图
$\boldsymbol{F}^m=\{\boldsymbol{f}_1^m,\boldsymbol{f}_2^m,\cdots\boldsymbol{f}_i^m,\cdots,\boldsymbol{f}_{N_i}^m\}$	多模态特征
$\boldsymbol{P}^m=\{\boldsymbol{p}_1^m,\boldsymbol{p}_2^m,\cdots,\boldsymbol{p}_u^m,\cdots,\boldsymbol{p}_{N_u}^m\}$	多模态偏好特征
$E=\{\boldsymbol{e}_{u_1},\cdots,\boldsymbol{e}_u,\boldsymbol{e}_{i_1},\cdots,\boldsymbol{e}_i\}$	初始嵌入集合
d	嵌入特征维度
D	多模态特征维度

2.3 层次锚点拓扑结构

本文通过层次锚点拓扑嵌入显式建模节点与全局结构关联性以增强节点特征,为去噪交互图生成器提供结构感知的节点表示,解决传统图神经网络难以捕获长距离依赖及过度平滑的问题。

2.3.1 层次锚点选择

具体而言,将用户节点集合 U 和项目节点集合 I 分别按节点度降序排列[28],依据度分布划分为三个活跃度层次,划分规则如图2所示。高活跃度层包含度值最高的前20%节点,主导交互图的信息传播与连通性;中活跃度层涵盖了度值处于中间40%的节点,涵盖普通用户与常规项目,构成交互结构的主体[29];低活跃度层则由度值较低的后40%节点组成,对应低频交互用户和长尾项目,用于丰富交互的多样性[30]。锚节点采样遵循6:3:1的比例从划分好的活跃度层次中随机采样锚节点 $V_A \subset V$,这种比例设置是基于对不同活跃度节点在交互图拓扑中的重要性和代表性的考量(在3.5章节进行论述)。其中为了保证用户和项目在全局拓扑表征中的作用,避免单侧信息主导拓扑学习,用户锚点和项目锚点的总数各占50%。

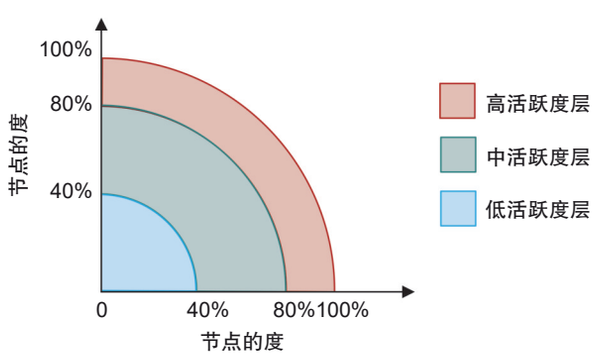

图2 层次锚点选择示意图

Fig. 2 Schematic diagram of hierarchical anchor selection

需要说明的是,若仅依赖初始交互频度进行静态划分,难以反映用户-项目交互的动态变化。为此,在迭代优化过程中,模型会基于最新的交互分布重新划分节点活跃度层次,并动态选择锚点节点,以保持用户与项目在全局拓扑表征中的均衡贡献。

2.3.2 层次锚点拓扑传播

基于选定的层次锚点,通过计算从节点 v_g 到节点 v_a 必须遍历的最小边数,来描述节点 v_g 与锚点 v_a 之间的距离 $d_{g,a}$,通过最短路径算法将衡量节点间在拓扑结构上的远近程度。拓扑权重 $\omega_{g,a}$ 由距离信息转化得到,如公式(1)所示:

$$\omega_{g,a}=\frac{1}{1+\exp(d_{g,a})}。\tag{1}$$

当节点与锚点的距离过大或没有连接时拓扑权重 $\omega_{g,a}$ 趋近于0,这表明当前节点与锚点间没有必要关联,不具备依据该锚点增强当前节点的必要性。为了增强用户、项目的节点特征,解决长距离依赖捕捉问题,层次锚点拓扑结构将依据相关权重来聚合节点嵌入,如公式(2)所示:

$$\bar{\boldsymbol{e}}_g^{(l)}=\sum_{(v_a\in V_A)}\boldsymbol{W}^{(l)}\cdot\omega_{g,a}\cdot\left[\bar{\boldsymbol{e}}_g^{(l-1)}\otimes\bar{\boldsymbol{e}}_a^{(l-1)}\right]\Big/|V_A|,\tag{2}$$

其中 $\boldsymbol{W}^{(l)}\in\mathbb{R}^{d\times 2d}$ 是第 l 传播层的权重矩阵,用于对节点特征进行线性变换;$\bar{\boldsymbol{e}}_g^{(l-1)},\bar{\boldsymbol{e}}_a^{(l-1)}\in\mathbb{R}^d$ 分别表示节点 v_g 和锚点 v_a 在第 $l-1$ 层的特征表示。通过将节点自身特征与锚点特征进行融合,可以为下文的去噪交互图生成器捕获全局结构的关联性。

2.4 多模态偏好辅助去噪交互图生成

为了抑制在挖掘用户潜在兴趣偏好过程中原始交互视图引入的噪声信息,设计了一种与多模态信息协同的去噪交互图生成器来学习用户偏好,并利用自监督学习策略抑制交互噪声。

2.4.1 用户-项目协同权重生成

本文采用多头注意力机制学习用户的真实交互偏好。为区分用户和项目所处不同的语义空间,分别定义不同的可学习矩阵 $\boldsymbol{Q}_u^h,\boldsymbol{W}_i^h,\boldsymbol{V}_i^h\in\mathbb{R}^{d/H}$,将增强后的用户特征作为查询张量,增强后的项目特征作为键张量和值张量,如公式(3)—公式(6)所示:

$$A_{ui}^h=(\boldsymbol{Q}_u^h\cdot\bar{\boldsymbol{e}}_u)^{\mathrm{T}}\cdot(\boldsymbol{W}_i^h\cdot\bar{\boldsymbol{e}}_i)\Big/\sqrt{d/H},\tag{3}$$

$$\overline{A}_{ui}^h=\mathrm{softmax}\left(A_{ui}^h\right),\tag{4}$$

$$\boldsymbol{Z}_u^H=\sum_{h=1}^{H}\overline{A}_{ui}^h\boldsymbol{V}_i^h\bar{\boldsymbol{e}}_i+\bar{\boldsymbol{e}}_u,\tag{5}$$

$$\boldsymbol{Z}_i^H=\sum_{h=1}^{H}\overline{A}_{ui}^h\boldsymbol{V}_i^h\bar{\boldsymbol{e}}_u+\bar{\boldsymbol{e}}_i,\tag{6}$$

其中 $\boldsymbol{Z}_u^H$ 和 $\boldsymbol{Z}_i^H$ 表示由多头注意力机制增强后的用户、项目嵌入表示集合。基于以上设计,模型可以学习到节点间的关联强度。将 $\overline{A}_{ui}^h$ 融合

来表示结构偏好权重α_{ui}^{h},该权重将作为协同权重$\bar{\alpha}_{ui}$的一部分影响交互图的去噪,该过程如公式(7)—公式(9)所示:

$$\bar{\alpha}_{ui}^{h}=\sum_{h=1}^{H}\overline{A}_{ui}^{h}/H\,, \tag{7}$$

$$\alpha_{ui}^{h}=\bar{\alpha}_{ui}^{h}/\sum_{(v_i,v_u)\in E}\bar{\alpha}_{ui}^{h}\,, \tag{8}$$

$$\bar{\alpha}_{ui}=\begin{cases}\alpha_{ui}^{h}+\psi\alpha_{ui}^{m}, & \text{存在模态偏好特征}\\ \alpha_{ui}^{h}, & \text{不存在模态偏好特征}\end{cases}, \tag{9}$$

其中α_{ui}^{m}为多模态偏好权重,ψ为多模态偏好权重的影响系数。

2.4.2 去噪交互图生成器

为了提升模型抑制交互噪声的能力,设计去噪交互图生成器,通过协同权重指导自动编码器实现交互图的自我强化。具体而言,模型运用协同权重分别对交互图G中高、低协同权重的边进行遮蔽形成噪声交互图G_V和增强交互图G_E。在此基础上,通过自动编码器进行交互重建。以G_V为例,其生成方法如公式(10)—公式(11)所示:

$$\alpha_{ui}^{V}=\mathrm{softmax}\left(\frac{1}{\bar{\alpha}_{ui}+\varepsilon}\right), \tag{10}$$

$$G_V=\mathrm{random}(V,\alpha_{ui}^{V}\bullet E)\,, \tag{11}$$

其中α_{ui}^{V}表示图G_V的边采样概率,ε为避免零分母而添加的小值。为了指导自动编码器通过自监督学习机制从噪声图中恢复高协同权重的交互,模型使用噪声图进行层次锚点拓扑嵌入操作,得到用户和项目的噪声嵌入特征$\boldsymbol{e}_u'$、$\boldsymbol{e}_i'$,同时构建损失函数$\mathcal{L}_{\mathrm{MPR}}$进行训练,使得噪声交互得分更低,如公式(12)所示:

$$\mathcal{L}_{\mathrm{MPR}}=\sum_{(v_u,v_i)\in G_V}\boldsymbol{e}'^{\mathrm{T}}_{u}\boldsymbol{e}_i'\,。 \tag{12}$$

为了增强模型对噪声的鲁棒性、减少信息冗余,避免已有研究中随机扰动视图对结构语义信息的破坏,本文采取基于自监督学习策略的图对比学习方法来过滤交互噪声。通过仅保留低协同权重交互生成对比交互图G_C,拉近同一用户或项目在增强交互图G_E中的表示特征与全局表征之间的距离,同时推远其在对比交互图中G_C的噪声交互表征。基于该优化过程,定义如公式(13)所示的最小化目标:

$$\mathcal{L}_{\mathrm{CLG}}=\log\sum_{v\in V}\exp\cos(\boldsymbol{e}_v^{E},\boldsymbol{e}_v^{C})/\tau\,, \tag{13}$$

其中$\boldsymbol{e}_v^{E}$和$\boldsymbol{e}_v^{C}$分别表示节点v在增强交互图和对比交互图中的高阶嵌入。去噪交互图生成器的最终损失如公式(14)所示:

$$\mathcal{L}_{\mathrm{DG}}=\mathcal{L}_{\mathrm{MPR}}+\lambda_0\bullet\mathcal{L}_{\mathrm{CLG}}\,, \tag{14}$$

其中λ_0为图结构对比学习损失权重。基于上述设计,由协同权重指导的自动编码器可以通过自监督学习实现交互图增强,从而抑制噪声对特征学习的影响。

2.5 增强交互图引导模态偏好生成

针对交互噪声与多模态信息利用不充分的问题,本文强化多模态信息在交互噪声分辨中的作用。通过建模用户跨模态偏好并将多模态信息融入噪声处理,挖掘用户与项目的深层特征关联,提升推荐系统在噪声环境中的鲁棒性与语义表征能力。

2.5.1 模态偏好生成器

为在挖掘项目多模态语义并建立用户跨模态偏好的同时规避交互噪声,模型基于增强交互图设计自注意力模块生成用户模态偏好特征。具体来说,模型将项目的多模态特征$\boldsymbol{F}^{m}$映射至增强交互图G_E上,采用LightGCN进行消息传播得到增强视图下的用户嵌入表示$\bar{\boldsymbol{e}}_u$。以音频模态为例,利用自注意力机制捕捉当前用户嵌入与项目在音频模态下的交互信息,如公式(15)—公式(17)所示:

$$\boldsymbol{Q}=\boldsymbol{W}^{Q}\bar{\boldsymbol{e}}_u,\boldsymbol{K}=\boldsymbol{W}^{K}\bar{\boldsymbol{e}}_u\,, \tag{15}$$

$$\bar{\boldsymbol{p}}_u^{A}=\mathrm{softmax}((\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}})/\sqrt{d}\,)\boldsymbol{W}^{V}\boldsymbol{f}_i^{A}\,, \tag{16}$$

$$\boldsymbol{p}_u^{A}=\mathrm{leakyrelu}(\boldsymbol{W}_A'\bullet\bar{\boldsymbol{p}}_u^{A}+\boldsymbol{b}_A')\,, \tag{17}$$

其中$\boldsymbol{W}^{Q},\boldsymbol{W}^{K},\boldsymbol{W}^{V}\in\mathbb{R}^{d\times d}$均为可学习的模型参数,$\boldsymbol{W}_A'\in\mathbb{R}^{d\times d}$和$\boldsymbol{b}_A'\in\mathbb{R}^{d}$分别为投影矩阵、偏置向量,$\boldsymbol{f}_i^{A}$为项目的音频特征。类似地,可以得到用户多模态偏好特征$\boldsymbol{p}_u^{A}$、$\boldsymbol{p}_u^{T}$和$\boldsymbol{p}_u^{V}$。

为了进一步提升用户多模态偏好特征的表征能力,缓解输入多模态语义信息带来的噪声问题,本文采用跨模态对比学习策略,将已交互项目的模态信息视为正样本对$(\boldsymbol{p}_u^{A},\boldsymbol{f}_{i^+}^{A})$,将未交互项目的模态信息视为负样本对$(\boldsymbol{p}_u^{A},\boldsymbol{f}_{i^-}^{A})$,通过损失函数$\mathcal{L}_{\mathrm{MCL}}$进行优化,如公式(18)所示:

$$\mathcal{L}_{\mathrm{MCL}}=-\sum_{u,i\in V}\log\frac{\exp\left(\mathrm{sim}\left(\boldsymbol{p}_u^{A},\boldsymbol{f}_{i^+}^{A}\right)/\tau\right)}{\sum\limits_{i\in i^+,i^-}\exp\left(\mathrm{sim}\left(\boldsymbol{p}_u^{A},\boldsymbol{f}_{i}^{A}\right)/\tau\right)}\,。 \tag{18}$$

2.5.2 模态偏好权重生成

本文通过生成各个模态下的偏好权重来完

成去噪交互图生成器与多模态偏好生成器之间的双向协同更新。采用余弦相似度计算多模态下节点间的语义关联,通过平均求和得到模态偏好权重α_{ui}^{m}。以音频模态为例,如公式(19)—公式(20)所示:

$$\alpha_{ui}^{A}=\mathrm{s}(\boldsymbol{p}_u^A,\boldsymbol{f}_i^A)=\frac{\boldsymbol{p}_u^A\cdot\boldsymbol{f}_i^A}{\|\boldsymbol{p}_u^A\|\|\boldsymbol{f}_i^A\|},\tag{19}$$

$$\alpha_{ui}^{m}=\sum_{M\in(A,V,T)}\alpha_{ui}^{M}/M。\tag{20}$$

这样的设计确保被保留下来的交互关系包含了所有模态的协同信息,能够更全面地刻画用户对项目的真实偏好。同时,设计了“图结构-模态”一致性约束,抑制模态噪声对图结构的污染,实现图结构与模态的语义对齐,如公式(21)所示:

$$\mathcal{L}_{\mathrm{GM}}=\sum_{(u,i)\in E}\|\alpha_{ui}^{m}-\alpha_{ui}^{h}\|^2。\tag{21}$$

2.6 模型优化与训练

对于用户u和项目i,将增强特征表示$\boldsymbol{Z}^H$基于增强交互图G_E进行传播,得到增强的高阶嵌入$\hat{\boldsymbol{E}}$。本文通过等权重平均融合高阶嵌入和各个模态下的用户偏好及项目多模态特征得到最终用户嵌入$\tilde{\boldsymbol{e}}_u$和项目嵌入$\tilde{\boldsymbol{e}}_i$,如公式(22)—公式(23)所示:

$$\hat{\boldsymbol{E}}=\mathrm{L}-\mathrm{GCN}^{L}(\boldsymbol{Z}^H,G_E),\tag{22}$$

$$\tilde{\boldsymbol{e}}_u=\hat{\boldsymbol{e}}_u+\beta_m\boldsymbol{p}_u^m,\tilde{\boldsymbol{e}}_i=\hat{\boldsymbol{e}}_i+\beta_m\boldsymbol{f}_i^m。\tag{23}$$

交互评分的计算如公式(24)所示。采用贝叶斯个性化排序(Bayesian Personalized Ranking, BPR)损失进行训练,如公式(25)所示:

$$\hat{y}_{ui}=\tilde{\boldsymbol{e}}_u^{\mathrm{T}}\tilde{\boldsymbol{e}}_i,\tag{24}$$

$$\mathcal{L}_{\mathrm{BPR}}=-\sum_{(u,i^+,i^-)\in O}\ln\sigma(\hat{y}_{ui^+}-\hat{y}_{ui^-}),\tag{25}$$

其中σ是sigmoid函数,O为训练集,i^+、i^-分别表示与用户u交互过和未交互过的项目,$\hat{y}_{ui^+}$、$\hat{y}_{ui^-}$为对应预测分数。模型通过优化手段高效建模用户-项目交互关系,并整合2.4节与2.5节提出的三种自监督学习损失函数,构建联合优化的总体损失函数,如公式(26)所示,以进一步挖掘图结构与节点关联信息,提升模型表征能力。

$$\begin{aligned}\mathcal{L}=&\mathcal{L}_{\mathrm{BPR}}+\mathcal{L}_{\mathrm{DG}}+\lambda_1\cdot\mathcal{L}_{\mathrm{MCL}}+\\&\lambda_2\cdot\mathcal{L}_{\mathrm{GM}}+\lambda_3\cdot\|\Theta\|_{\mathrm{F}}^2,\end{aligned}\tag{26}$$

其中λ_1、λ_2和λ_3是各优化任务的权重。Θ表示模型的参数及其正则化项。

3 实验及分析

为了验证HEMMR的有效性,本文针对如下问题进行了实验。

问题1:与经典的协同过滤推荐模型和最新的多模态推荐模型基线相比,HEMMR性能如何。

问题2:各模块的工作对HEMMR的性能有什么影响。

问题3:超参数对HEMMR模型的性能有何影响。取何值时模型性能达到最优。

问题4:层次锚点拓扑结构中锚点采样比例是否具备泛化能力,采用动态的可学习采样比例是否更好。

问题5:模态和图结构双向协同去噪和自监督学习训练算法是否帮助HEMMR模型提高了在噪声环境下的鲁棒性。

问题6:通过层次锚点拓扑结构是否能有效解决过度平滑和捕捉长距离依赖困难的问题。

3.1 实验设置

为验证文中提出的HEMMR的性能,本文基于3个公开的多模态推荐数据集TikTok、Baby和Sports开展实验,数据集的统计信息见表2。

表2 数据集统计信息

Table 2 Statistics of datasets

名称	用户数量	产品数量	交互数量	稀疏度/%	模态	特征维度
TikTok	9 308	6 710	68 722	99.89	视觉	128
					文本	768
					音频	128
Baby	19 445	7 050	160 792	99.88	视觉	4 096
					文本	1 024
Sports	18 164	14 514	119 708	99.95	视觉	4 096
					文本	1 024

实验时选择PyTorch为机器学习的框架,在PyCharm 2021, Intel®CoreIntel(R) Core(TM) i7-8700K CPU@3.70 GHz, Python 3.7, Win10 64位操作系统的实验环境下进行对比实验。在实验中,节点嵌入维度d确定为64,Batch-size设定为1024。采用Adam优化算法,设定初始学习率为0.001。选取Recall@20、NDCG@20作为

评价指标,通过三次重复实验取其平均值,以确保数据可靠性与结论稳健性。

3.2 对比实验

针对问题1,本文将HEMMR与多类基线模型对比,包括传统推荐模型(MF-BPR[4]),自监督学习模型(SGL[19]、SimGCL[20]、GFormer[21]),图神经网络推荐模型(LightGCN[12]、NGCF[6]),多模态推荐系统(MGCN[16]、LATTICE[14]、BM3[23]、LGMRec[26]、MMGCL[22]、PGL[27]、SLMRec[24])。表3实验结果显示,HEMMR在3个数据集的Recall@20、NDCG@20指标上均优于所有基线模型。

相较于传统推荐方法MF-BPR,图神经网络模型NGCF和LightGCN凭借高阶协同表征进行推荐在推荐性能上实现了突破。其中LightGCN因架构精简更具优势,但在长距离依赖捕捉问题上表现不足。自监督模型中,SGL和SimGCL等模型均在图神经网络中整合了对比学习范式。实验数据表明,而SGL的随机视图生成方法因忽略语义关联,导致性能受限。GFormer因自监督学习任务与真实交互存在一定的语义偏差,所以在稀疏场景表现较弱。

表3 HEMMR与各对比模型性能比较

Table 3 Performance comparison between HEMMR and comparative models

模型	TikTok		Baby		Sports	
	Recall@20	NDCG@20	Recall@20	NDCG@20	Recall@20	NDCG@20
MF-BPR	0.041 3	0.021 1	0.043 4	0.020 1	0.041 2	0.020 0
LightGCN	0.072 1	0.036 9	0.068 9	0.030 1	0.069 0	0.031 2
NGCF	0.068 4	0.028 4	0.058 2	0.026 0	0.061 8	0.030 9
SGL	0.067 8	0.032 6	0.067 1	0.029 6	0.060 6	0.028 7
SimGCL	0.086 8	0.035 9	0.067 4	0.029 3	0.071 3	0.032 3
GFormer	0.079 4	0.034 3	0.071 0	0.032 1	0.071 9	0.033 1
MGCN	0.096 5	0.041 7	0.090 9	0.038 9	0.091 3	0.040 3
LATTICE	0.090 4	0.039 5	0.085 1	0.036 6	0.087 0	0.039 4
BM3	0.094 7	0.040 2	0.082 4	0.034 8	0.083 1	0.039 1
LGMRec	0.085 4	0.035 0	<u>0.091 5</u>	<u>0.040 1</u>	<u>0.092 1</u>	<u>0.042 1</u>
MMGCL	0.081 3	0.032 8	0.076 6	0.034 2	0.085 6	0.040 1
PGL	0.085 1	0.034 6	0.088 7	0.038 8	0.083 4	0.036 4
SLMRec	<u>0.102 7</u>	<u>0.042 3</u>	0.080 5	0.036 1	0.085 6	0.038 6
HEMMR	**0.104 2**	**0.043 8**	**0.093 1**	**0.040 8**	**0.095 7**	**0.044 0**
提升	1.46%	3.54%	1.75%	1.75%	3.91%	4.51%

注:Recall@20表示模型推荐的前20个项目中,命中用户真实交互过的项目的比例;NDCG@20表示推荐的前20个项目的排序质量。最好的基线模型效果采用下划线标出,最强的效果采用粗体标出。

相比上述模型,多模态推荐方法在基准模型中展现出系统性优势。图神经网络多模态方法MGCN通过融合模态与图结构展现优势,验证了多模态信息的关键价值。引入对比学习的模型LATTICE、BM3、MMGCL通过自监督任务增强模态表征,推动性能提升。SLMRec借助多模态信息辅助模型感知用户偏好,通过特征丢弃、特征掩蔽等操作生成单个项目的多个视图,但未对原始交互图结构进行去噪,这不利于挖掘用户和项目间关系。PGL通过挖掘和利用主要的局部结构来增强用户偏好,实现性能提升。LGMRec同时对局部和全局信息建模,分别利用局部图和全局超图捕获局部拓扑关系和全局依赖,又进一步提升模型性能。

实验结果表明简单的对比学习范式和基于协同信息直接建模用户偏好的推荐模型未能充分释放多模态潜力。而本文的HEMMR通过层次锚点拓扑结构解决长距离依赖捕捉问题,并通过多头注意力机制充分利用多模态特征参与交互去噪,同时利用“图结构-模态”一致性约束及对比学习抑制噪声,增强特征表示能力。其基于多模态偏好的交互图去噪机制与自监督协同训练策略,能有效区分真实交互与噪声、语义关联与模态噪声,最终实现性能突破。

与此同时,为了验证HEMMR的训练效率,本文在TikTok和Sports数据集上与SGL、SLMRec、BM3、LGMRec等四种代表性模型进行了收敛速度对比实验,结果如图3所示。HEMMR的优势主要来自三方面:其层次锚点拓扑结构增强了节点的全局结构感知,加快了长距离依赖的学习;可学习生成器协同建模多模态偏好和去噪交互图,有效降低了初期噪声干扰;图结构-模态一致性约束实现了信息双向对齐,进一步提升了优化效率。相比之下,SGL依赖随机视图增强,缺乏结构引导,优化过程缓慢;SLMRec和BM3分别采用静态图或随机扰动生成对比视图,难以早期捕捉真实偏好;LGMRec虽结合局部和全局信息建模,但其采用超图神经网络捕获更复杂的关系,这使得传播路径长、冗余信息多,导致训练过程收敛较慢。

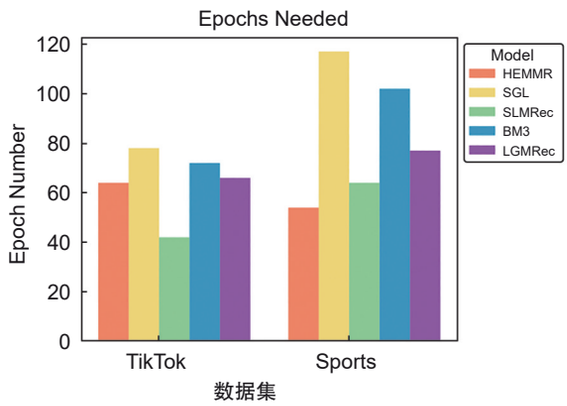

图3 方法的收敛速度对比

Fig. 3 Comparison of convergence speed of methods

3.3 消融实验

针对问题2,本文设计了5种变体模型开展消融实验,从层次锚点拓扑嵌入、多模态偏好辅助交互图去噪、去噪交互图引导多模态偏好生成及"图结构-模态"一致性四个维度进行分析,具体变体模型包括:

1)w/o HAS(Hierarchical Anchor Selection):以随机锚点代替层次锚点选择。

2)w/o HAT(Hierarchical Anchor Topology):移除层次锚点拓扑嵌入结构。

3)w/o GMC(Graph-modal Consistency):删除"图结构-模态"一致性损失。

4)w/o GG(Graph-guided):使用原交互结构代替去噪交互图引导模态偏好生成。

5)w/o MG(Modality-guided):取消多模态偏好辅助去噪交互图生成器。

消融实验结果如表4所示。w/o HAS在保留拓扑结构的基础上随机选用锚点,不考虑节点的活跃度对拓扑结构的影响,使得HEMMR不能准确捕获流行度偏差对节点拓扑的影响,从而无法解决长距离依赖捕捉问题。而w/o HAT禁用拓扑结构的表现较差是由于变体模型无法捕获全局节点的分布情况,这导致去噪交互图生成器缺乏对全局协同关系的感知,让模型只能依赖稀疏的局部邻居判断交互噪声,因此性能下降。

表4 消融实验结果

Table 4 Results of ablation study

变体模型	TikTok		Baby		Sports	
	Recall@20	NDCG@20	Recall@20	NDCG@20	Recall@20	NDCG@20
w/o HAS	0.096 1	0.041 1	0.088 5	0.038	0.090 9	0.041 0
w/o HAT	0.093 6	0.037 4	0.084 4	0.033 2	0.091 5	0.041 8
w/o GMC	0.102 0	0.041 8	0.089 8	0.038 1	0.092 4	0.042 3
w/o GG	0.100 2	0.042 7	0.088 7	0.037 5	0.090 8	0.041 5
w/o MG	0.097 4	0.041 7	0.089 6	0.039 0	0.087 1	0.040 5
HEMMR	**0.104 2**	**0.043 8**	**0.093 1**	**0.040 8**	**0.095 7**	**0.044 0**

注:消融实验用于验证HEMMR各核心模块的必要性,Recall@20和NDCG@20值越大性能越优,最强的效果采用粗体标出。

w/o GG不使用去噪后的交互图生成用户模态偏好,实验结果说明仅依靠原始交互无法规避噪声交互的影响,因此难以挖掘用户的真实模态偏好。而w/o MG未使用多模态信息进行交互噪声的判别,将导致去噪交互图生成器无法考虑多模态特征中的协同信号,致使性能下降。值得注意的是,w/o MG变体在TikTok数据集上有大幅下降。这是因为该数据集模态信息较多,运用多模态信息能更好地捕获用户的真实偏好,从而完成对交互关系的去噪,提

升算法性能。

w/o GMC 不采用“图结构-模态”一致性约束,这使得 HEMMR 在面临结构偏好权重与模态偏好权重呈相反态势时无法准确判别交互噪声。由于 Sports 数据集数据较稀疏,这一问题表现得尤为显著。

3.4 参数敏感性分析

针对问题 3,本节对 HEMMR 的 4 个关键超参数进行讨论:图结构对比学习损失权重 λ_0,模态间对比学习损失权重 λ_1,“图结构-模态”一致性约束损失函数权重 λ_2 和模态偏好权重在协同权重中的占比超参数 ψ。通过网格搜索方法在三个数据集上寻找最优值。

由图 4 可得对比学习任务对模型性能存在显著影响。值得注意的是,Baby 数据集对高对比权重更为敏感。这是因为其项目空间远小于用户空间,这种节点数量的不匹配将导致稀疏交互带来的噪声影响进一步放大,因此需放大对比学习任务的影响以增强表征。

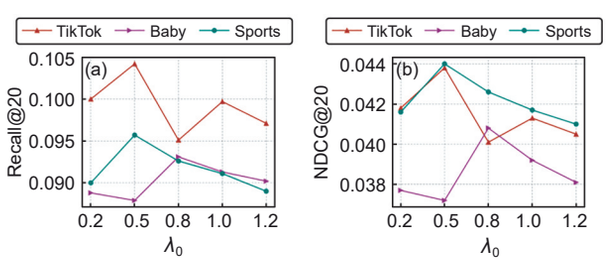

图 4 图结构对比学习损失权重 λ_0 对模型性能的影响

(a)不同 λ_0 取值下 Recall@20 变化趋势;(b)不同 λ_0 取值下 NDCG@20 变化趋势

Fig. 4 Impact of graph structure contrastive learning loss weight λ_0 on performance of model

(a) Variation trends of Recall@20 under different λ_0 values; (b) variation trends of NDCG@20 under different λ_0 values

参数 λ_1 影响如图 5 所示。TikTok 作为短视频推荐场景,模态信息质量较高,过大的 λ_1 值会导致模态一致性升高从而抹除特定模态信息对用户偏好的影响。相反,Baby 和 Sports 作为经典的电商推荐场景,海量的交互项目会导致用户的历史记录稀疏度较高,因此模型需要较大的 λ_1 值来增强跨模态一致性,抵消多模态噪声带来的影响。

λ_2 和 ψ 在三个数据集上的具体影响如图 6—图 7 所示。实验结果与前文分析一致,在电商场景下模态语义可能有较大差异,需要更高的“图结构-模态”一致性约束 λ_2 来抑制噪声,而模态偏好权重在协同权重中的占比 ψ 则可能因多模态信息对交互影响较小而取较低值。TikTok 数据集则因模态信息丰富且差异性较大,实验结果表明强约束会抑制有效的模态分歧信息。

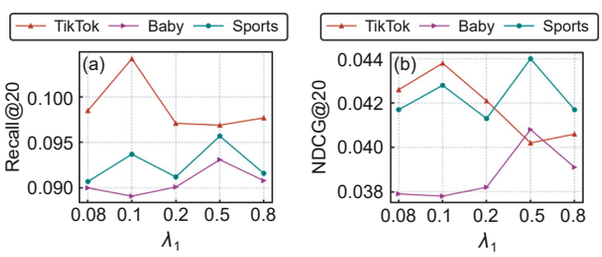

图 5 模态间对比学习损失权重 λ_1 对模型性能的影响

(a)不同 λ_1 取值下 Recall@20 变化趋势;(b)不同 λ_1 取值下 NDCG@20 变化趋势

Fig. 5 Impact of inter-modal contrastive learning loss weight λ_1 on performance of model

(a) Variation trends of Recall@20 under different λ_1 values; (b) variation trends of NDCG@20 under different λ_1 values

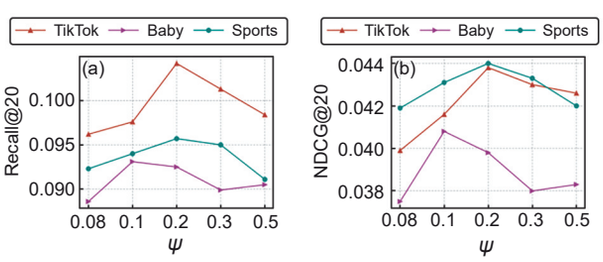

图 6 协同权重中的占比超参数 ψ 对模型性能的影响

(a)不同 ψ 取值下 Recall@20 变化趋势;(b)不同 ψ 取值下 NDCG@20 变化趋势

Fig. 6 Impact of proportion hyperparameter in collaborative weight ψ on performance of model

(a) Variation trends of Recall@20 under different ψ values; (b) variation trends of NDCG@20 under different ψ values

3.5 锚点采样比例泛化能力研究

针对问题 4,为验证层次锚点拓扑结构中锚点采样比例的合理性与泛化能力,实验对比了动态可学习采样比例及三种静态比例(3∶4∶3、2∶4∶4、8∶1∶1)在 TikTok 和 Sports 数据集上的推荐性能。结果显示如表 5 所示,原设定 6∶3∶1 在 Recall@20 和 NDCG@20 指标上均取得最优表现。

动态可学习采样比例实验是通过在算法引入一个三维可学习参数向量来表示各层次节点的选择比例,同时通过 Softmax 函数约束,确保

选择比例的有效性。在训练过程中,损失函数的梯度会更新参数向量,该参数向量通过反向传播机制进行端到端优化。然而该学习驱动的动态采样比例在Recall@20和NDCG@20上未能超越静态设置的6:3:1比例。这样的实验结果是由于在模型HEMMR中,动态采样机制易在训练中因高频锚点提供的信号更快产生性能提升而向高频采样偏移,破坏全局建模能力,且增加训练复杂度。

观察其他的锚点采样比例实验可以发现过多引入低频锚点可能导致噪声干扰和连通性下降,而高频锚点比例过高则会引发过度平滑问题。HEMMR设计6:3:1的采样比例在高活跃度、中活跃度与低活跃度节点间构建了良好的结构平衡,能够同时捕捉高频主干传播路径与低频长尾节点的多样性特征,使得模型能够捕获全局结构的关联性。

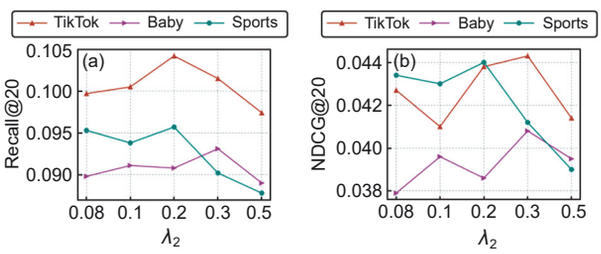

图7 "图结构-模态"一致性约束损失函数权重λ_2对模型性能的影响

(a)不同λ_2取值下Recall@20变化趋势;(b)不同λ_2取值下NDCG@20变化趋势

Fig. 7 Impact of the "graph-structure-modality" consistency constraint loss function weight λ_2 on performance of model

(a) Variation trends of Recall@20 under different λ_2 values; (b) variation trends of NDCG@20 under different λ_2 values

表5 多种锚点采样比例对比研究

Table 5 Comparison of multiple anchor sampling ratios

比例设定	TikTok		Sports	
	Recall@20	NDCG@20	Recall@20	NDCG@20
3:4:3	0.100 7	0.042 5	0.088 7	0.040 2
2:4:4	0.100 3	0.042 0	0.088 4	0.040 3
8:1:1	0.100 4	0.042 1	0.089 2	0.040 3
动态可学习采样	0.100 0	0.042 2	0.090 2	0.042 1
6:3:1	**0.104 2**	**0.043 8**	**0.095 7**	**0.044 0**

注:所有指标保留至小数点后四位数,Recall@20和NDCG@20值越大性能越优,最强的效果采用粗体标出。

3.6 噪声的鲁棒性研究

针对问题5,为评估HEMMR对交互噪声的鲁棒性,本文以Baby数据集为基础,通过注入10%~50%的随机噪声交互观察模型性能的衰减情况。选择SGL和LGMRec作为对比的基线模型,实验结果如图8所示。

结果表明,HEMMR在高噪声场景下的衰减率显著低于基线模型。其中,SGL的实验结果显示传统的随机扰动构造增强视图的方式会导致难以区分噪声边与真实交互。而LGMRec虽依赖模态特征构造了局部与全局视图的对比学习,但忽视了图结构固有的噪声问题。这导致在噪声增加时,其图结构被污染。所以进行模态传播过程时会再次引入噪声,导致性能下降较快。HEMMR利用多模态信息来辅助生成去噪交互图,通过多头注意力机制,模型能够从多模态信息中提取用户真实偏好,从而更新交互权重。凭借模态和图结构双向协同去噪和自监督学习训练算法,HEMMR在噪声环境下凸显价值。

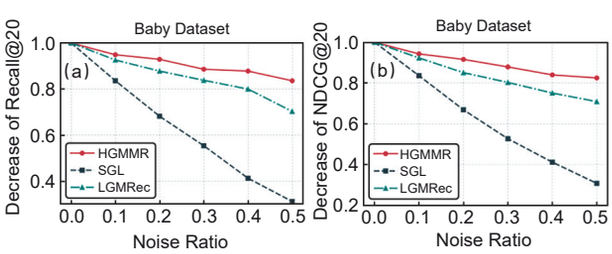

图8 Baby数据集上的噪声扰动下的鲁棒性

(a)噪声环境下模型性能Recall@20的变化趋势;(b)噪声环境下模型性能NDCG@20的变化趋势

Fig. 8 Performance under noise perturbation on the Baby dataset

(a) Variation trends of model performance in terms of Recall@20 under noisy environment; (b) variation trends of model performance in terms of NDCG@20 under noisy environment

3.7 特征分布研究

针对问题6,为评估HEMMR的特征表示能力设计了如图9所示的可视化实验。实验

选择基线模型LGMRec作为对比,在Sports数据集上随机采样1 000个项目并使用t-SNE绘制2D特征分布的散点图。图底部的高斯核密度估计显示了节点嵌入在单位超球面S^1上的密度分布。从图中可以看出,LGMRec的特征分布呈现较为明显的聚类簇,这是由于堆叠多层图卷积和超图卷积引起的过度平滑问题会使得节点表征趋于相似,因此聚集在一起。体现在密度图上为部分方向上存在波峰分布,反映出节点表征在空间中聚集在少数特定方向上,因此模型未能学习到多样化的项目特征。

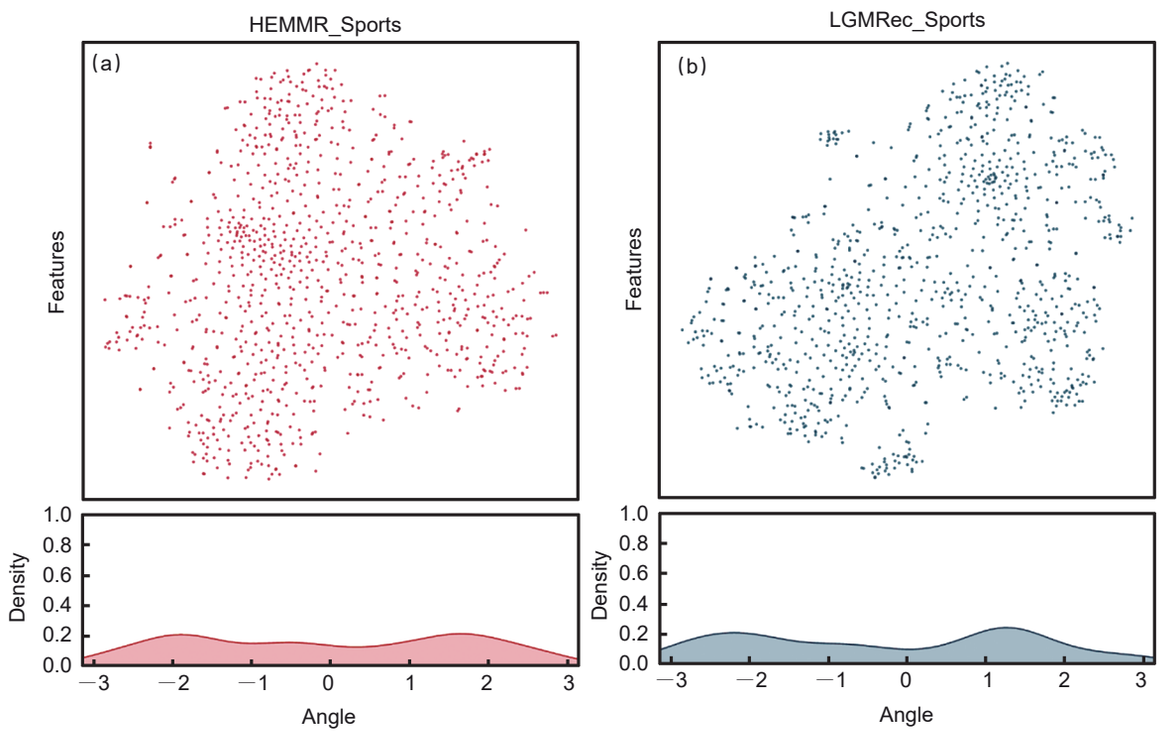

图9 Sports数据集上的特征分布

(a) HEMMR在Sports数据集上的特征分布;(b) LGMRec在Sports数据集上的特征分布

Fig. 9 Feature distribution on the Sports dataset

(a) Feature distribution of HEMMR on the Sports dataset; (b) feature distribution of LGMRec on the Sports dataset

相较之下,HEMMR通过层次锚点拓扑结构增强用户与项目节点特征,有效捕捉不同层次的特征信息,从而帮助模型学习全局关联性,其特征在坐标平面内分布更为均匀。此外,密度曲线相对平缓,各轴值区间密度分布均匀,表明模型缓解了GCN带来的过度平滑问题,并进一步验证其充分利用长距离依赖信息来提升推荐质量的优势。

4 结论

本文针对推荐系统中多模态信息利用不足与交互数据噪声干扰的问题,提出了一种层次拓扑增强的多模态推荐方法,并在多个真实数据集上进行了系统实验。通过层次化锚点选择,模型能够更全面地捕捉用户与项目在不同活跃度下的全局拓扑特征,使节点表征保持多样性与区分性。本文设计的生成器能够同时重建用户多模态偏好与去噪交互图,并借助“图结构-模态”一致性约束实现双向协同更新。实验结果显示,引入该机制能有效缓解随机噪声和复杂交互带来的性能衰减,尤其在噪声水平较高的场景中优势更为突出。

尽管本文提出的方法在准确性与鲁棒性上均取得了良好表现,但仍存在改进空间。多项损失函数的加权组合可能在更复杂场景下引发梯度冲突与收敛不一致,未来可结合动态权重学习与多任务优化框架进一步提升训练稳定性。

参考文献:

[1] Kumar C, Chowdary C R, Meena A K. Recent Trends in

Recommender Systems: A Survey[J]. *Int J Multimed Inf Retr*, 2024, **13**(4): 41. DOI: 10.1007/s13735-024-00349-1.

[2] Ullah F, Zhang B F, Zou G B, *et al*. Large-scale Distributive Matrix Collaborative Filtering for Recommender System[C]//Proceedings of the 2020 International Conference on Computing, Networks and Internet of Things. New York: ACM, 2020: 55−59.. DOI: 10.1145/3398329.3398360.

[3] Hasan M, Nalagandla R. An Ensemble Weighted User-based Collaborative Filtering Recommender System[C]// 2024 2nd International Conference on Artificial Intelligence, Blockchain, and Internet of Things (AIBThings). New York: IEEE, 2024: 1−6. DOI: 10.1109/AIBThings63359.2024.10863411.

[4] Rendle S, Freudenthaler C, Gantner Z, *et al*. BPR: Bayesian Personalized Ranking from Implicit Feedback[C]//Proceedings of the Twenty−Fifth Conference on Uncertainty in Artificial Intelligence. Arlington, Virginia, USA: AUAI Press, 2009: 452−461. DOI: 10.5555/1795114.1795167.

[5] Jiang B, Zhang Z Y, Lin D D, *et al*. Semi-supervised Learning with Graph Learning-convolutional Networks[C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2019: 11305−11312. DOI: 10.1109/CVPR.2019.01157.

[6] Wang X, He X N, Wang M, *et al*. Neural Graph Collaborative Filtering[C]//Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2019: 165−174.. DOI: 10.1145/3331184.3331267.

[7] Zhou X, Lin D H, Liu Y, *et al*. Layer-refined Graph Convolutional Networks for Recommendation[C]//2023 IEEE 39th International Conference on Data Engineering (ICDE). New York: IEEE, 2023: 1247−1259. DOI: 10.1109/ICDE55515.2023.00100.

[8] Zhou P, Cui Y C, Cao H. A Power Method to Alleviate Over-smoothing for Recommendation[C]//Proceedings of the 33rd ACM International Conference on Information and Knowledge Management. New York: ACM, 2024: 3484−3493.. DOI: 10.1145/3627673.3679553.

[9] Yu J L, Yin H Z, Xia X, *et al*. Self-supervised Learning for Recommender Systems: a Survey[J]. *IEEE Trans Knowl Data Eng*, 2024, **36**(1): 335−355. DOI: 10.1109/TKDE.2023.3282907.

[10] Zheng Y P, Yi L, Wei Z W. A Survey of Dynamic Graph Neural Networks[J]. *Front Comput Sci*, 2024, **19**(6): 196323. DOI: 10.1007/s11704-024-3853-2.

[11] Liu Q D, Hu J X, Xiao Y T, *et al*. Multimodal Recommender Systems: A Survey[J]. *ACM Comput Surv*, 2025, **57**(2): 1−17. DOI: 10.1145/3695461.

[12] He X N, Deng K, Wang X, *et al*. LightGCN: Simplifying and Powering Graph Convolution Network for Recommendation[C]//Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2020: 639−648.. DOI: 10.1145/3397271.3401063.

[13] Wei Y W, Wang X, Nie L Q, *et al*. MMGCN: Multi-modal Graph Convolution Network for Personalized Recommendation of Micro-video[C]//Proceedings of the 27th ACM International Conference on Multimedia. New York: ACM, 2019: 1437−1445.. DOI: 10.1145/3343031.3351034.

[14] Zhang J H, Zhu Y Q, Liu Q, *et al*. Mining Latent Structures for Multimedia Recommendation[C]//Proceedings of the 29th ACM International Conference on Multimedia. New York: ACM, 2021: 3872−3880. DOI: 10.1145/3474085.3475259.

[15] Wei Y W, Liu W Q, Liu F, *et al*. LightGT: A Light Graph Transformer for Multimedia Recommendation [C]//Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2023: 1508−1517.. DOI: 10.1145/3539618.3591716.

[16] Yu P H, Tan Z Y, Lu G M, *et al*. Multi-view Graph Convolutional Network for Multimedia Recommendation [C]//Proceedings of the 31st ACM International Conference on Multimedia. New York: ACM, 2023: 6576−6585. DOI: 10.1145/3581783.3613915.

[17] Ren X B, Xia L H, Yang Y H, *et al*. SSLRec: A Self-supervised Learning Framework for Recommendation [C]//Proceedings of the 17th ACM International Conference on Web Search and Data Mining. New York: ACM, 2024: 567−575. DOI: 10.1145/3616855.3635814.

[18] Gheewala S, Xu S X, Yeom S. In-depth Survey: Deep Learning in Recommender Systems: Exploring Prediction and Ranking Models, Datasets, Feature Analysis, and Emerging Trends[J]. *Neural Comput Appl*, 2025, **37**(17): 10875−10947. DOI: 10.1007/s00521-024-10866-z.

[19] Wu J C, Wang X, Feng F L, *et al*. Self-supervised Graph Learning for Recommendation[C]//Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2021: 726−735. DOI: 10.1145/3404835.3462862.

[20] Yu J L, Yin H Z, Xia X, *et al*. Are Graph Augmentations Necessary?: Simple Graph Contrastive Learning for Recommendation[C]//Proceedings of the 45th Interna-

tional ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2022: 1294−1303.. DOI: 10.1145/3477495.3531937.

[21] Li C L, Xia L H, Ren X B, *et al*. Graph Transformer for Recommendation[C]//Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2023: 1680−1689.. DOI: 10.1145/3539618.3591723.

[22] Yi Z X, Wang X, Ounis I, *et al*. Multi-modal Graph Contrastive Learning for Micro-video Recommendation [C]//Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2022: 1807−1811.. DOI: 10.1145/3477495.3532027.

[23] Zhou X, Zhou H Y, Liu Y, *et al*. Bootstrap Latent Representations for Multi-modal Recommendation[C]//Proceedings of the ACM Web Conference 2023. New York: ACM, 2023: 845−854.. DOI: 10.1145/3543507.3583251.

[24] Tao Z L, Liu X H, Xia Y W, *et al*. Self-supervised Learning for Multimedia Recommendation[J]. *IEEE Trans Multimed*, 2022, **25**: 5107−5116. DOI: 10.1109/TMM.2022.3187556.

[25] Wei W, Huang C, Xia L H, *et al*. Multi-modal Self-supervised Learning for Recommendation[C]//Proceedings of the ACM Web Conference 2023. New York: ACM, 2023: 790−800.. DOI: 10.1145/3543507.3583206.

[26] Guo Z Q, Li J J, Li G H, *et al*. LGMRec: Local and Global Graph Learning for Multimodal Recommendation[J]. *Proc AAAI Conf Artif Intell*, 2024, **38**(8): 8454−8462. DOI: 10.1609/aaai.v38i8.28688.

[27] Yu P H, Tan Z Y, Lu G M, *et al*. Mind Individual Information! Principal Graph Learning for Multimedia Recommendation[J]. *Proc AAAI Conf Artif Intell*, 2025, **39**(12): 13096−13105. DOI: 10.1609/aaai.v39i12.33429.

[28] Vaez Barenji S, Parajuli S, Ekstrand M D. User and Recommender Behavior over Time: Contextualizing Activity Effectiveness Diversity and Fairness in Book Recommendation[C]//Adjunct Proceedings of the 33rd ACM Conference on User Modeling, Adaptation and Personalization. New York: ACM, 2025: 280−287.. DOI: 10.1145/3708319.3733710.

[29] Chen J W, Dong H D, Wang X, *et al*. Bias and Debias in Recommender System: A Survey and Future Directions[J]. *ACM Trans Inf Syst*, 2023, **41**(3): 1−39. DOI: 10.1145/3564284.

[30] Lin Z H, Tian C X, Hou Y P, *et al*. Improving Graph Collaborative Filtering with Neighborhood-enriched Contrastive Learning[C]//Proceedings of the ACM Web Conference 2022. New York: ACM, 2022: 2320−2329.. DOI: 10.1145/3485447.3512104.