



融合扩散模型的多模态图对比学习推荐

沈学利, 于佳灏

引用本文:

沈学利, 于佳灏. 融合扩散模型的多模态图对比学习推荐[J]. 智能系统学报, 2026, 21(4): 1044–1054.

SHEN Xueli, YU Jiahao. Multimodal graph contrastive learning recommendation based on fusion diffusion models[J]. *CAAI Transactions on Intelligent Systems*, 2026, 21(4): 1044–1054.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202507015>

您可能感兴趣的其他文章

融入学习者模型在线学习资源协同过滤推荐方法

A collaborative filtering recommendation method for online learning resources incorporating the learner model

智能系统学报. 2021, 16(6): 1117–1125 <https://dx.doi.org/10.11992/tis.202009005>

基于知识图谱和用户长短期偏好的个性化景点推荐

Personalized attraction recommendation based on the knowledge graph and users' long-term and short-term preferences

智能系统学报. 2020, 15(5): 990–997 <https://dx.doi.org/10.11992/tis.201904064>

面向自动驾驶目标检测的深度多模态融合技术

Deep multi-modal fusion in object detection for autonomous driving

智能系统学报. 2020, 15(4): 758–771 <https://dx.doi.org/10.11992/tis.202002010>

旅游知识图谱特征学习的景点推荐

Tourism knowledge-graph feature learning for attraction recommendations

智能系统学报. 2019, 14(3): 430–437 <https://dx.doi.org/10.11992/tis.201810032>

知识图谱的推荐系统综述

Review of recommendation systems based on knowledge graph

智能系统学报. 2019, 14(2): 207–216 <https://dx.doi.org/10.11992/tis.201805001>

结合稀疏表示与约束传递的半监督谱聚类算法

A semi-supervised spectral clustering algorithm combined with sparse representation and constraint propagation

智能系统学报. 2018, 13(5): 855–863 <https://dx.doi.org/10.11992/tis.201703013>

DOI: 10.11992/tis.202507015

网络出版地址: <https://link.cnki.net/urlid/23.1538.tp.20260403.1038.006>

融合扩散模型的多模态图对比学习推荐

沈学利, 于佳灏

(辽宁工程技术大学软件学院, 辽宁 葫芦岛 125105)

摘要: 针对多模态数据中固有噪声干扰、数据稀疏性以及跨模态语义表示差异导致的用户兴趣偏好建模不准确的问题, 提出了一种融合扩散机制的图对比学习推荐方法 (diffusion-enhanced exploration for multimodal graph contrastive recommendation, DEMGR)。借助扩散模型生成结构增强的对比视图, 有效抑制自监督学习中各个模态噪声的干扰; 构建融合用户与物品的同构图结构, 增强图的连通性, 缓解稀疏性对图表示学习的影响; 设计一个联合优化框架, 融合多任务学习与扩散增强机制, 协同优化推荐目标与增强信号, 学习具有语义一致性与泛化能力的多模态表示。在 Baby、Sports 和 Clothing 共 3 个真实数据集上的实证研究表明, DEMGR 在 Recall 和 NDCG(normalized discounted cumulative gain) 等指标上均显著优于现有主流模型, 最大提升分别达 5.45% 和 5.01%, 充分验证了该方法在多模态推荐中的有效性。本文分析结果可为多模态推荐系统优化提供参考。

关键词: 推荐系统; 扩散模型; 多模态; 图对比学习; 自监督学习; 多任务学习; 数据稀疏; 用户兴趣

中图分类号: TP311 **文献标志码:** A **文章编号:** 1673-4785(2026)04-1044-11

中文引用格式: 沈学利, 于佳灏. 融合扩散模型的多模态图对比学习推荐 [J]. 智能系统学报, 2026, 21(4): 1044-1054.

英文引用格式: SHEN Xueli, YU Jiahao. Multimodal graph contrastive learning recommendation based on fusion diffusion models[J]. CAAI transactions on intelligent systems, 2026, 21(4): 1044-1054.

Multimodal graph contrastive learning recommendation based on fusion diffusion models

SHEN Xueli, YU Jiahao

(School of Software, Liaoning Technical University, Huludao 125105, China)

Abstract: To address the inaccurate modeling of user interest preferences caused by noise interference, data sparsity, and cross-modal semantic discrepancies in multimodal data, this paper proposes a diffusion-enhanced exploration framework for multimodal graph contrastive recommendation(DEMGR). First, a diffusion model is employed to generate structurally enhanced contrastive views, effectively suppressing modality-specific noise during self-supervised learning. Second, an isomorphic graph integrating users and items is constructed to strengthen graph connectivity and mitigate the effects of data sparsity on graph representation learning. Finally, a joint optimization framework is designed to combine multitask learning with diffusion enhancement, enabling the collaborative optimization of recommendation objectives and enhancement signals while learning semantically consistent and generalizable multimodal representations. Experiments conducted on three real-world datasets, Baby, Sports, and Clothing, demonstrate that DEMGR consistently outperforms existing mainstream models in terms of Recall and normalized discounted cumulative gain(NDCG), achieving maximum improvements of 5.45% and 5.01%, respectively. These results verify the effectiveness of the proposed method for multimodal recommendation and provide valuable insights for the optimization of multimodal recommender systems.

Keywords: recommender system; diffusion model; multimodal; graph contrastive learning; self-supervised learning; multitask learning; data sparsity; user interest

随着信息时代的发展, 推荐系统已成为缓解

信息过载、提升用户体验的关键技术。传统推荐方法多基于用户历史交互数据^[1], 难以全面刻画用户兴趣和物品特性。多模态推荐系统 (multimodal re-commendation systems, MRS) 通过整合

收稿日期: 2025-07-11. 网络出版日期: 2026-04-03.

基金项目: 国家自然科学基金项目 (62173171).

通信作者: 沈学利. E-mail: lngd011@163.com.

©《智能系统学报》编辑部版权所有

文本、图像等模态信息,构建更丰富的用户与物品表示,在电商^[2]与短视频平台^[3]等应用中展现出强大潜力。然而,高效融合异构模态数据、发挥模态间协同效应仍是挑战。早期如 VBPR(visual bayesian personalized ranking)^[4]将视觉特征引入矩阵分解,融合方式较浅;后续 DualGNN(dual graph neural network)^[5]利用图神经网络建模多模态特征,并引入用户共现图增强偏好挖掘,尽管性能提升,但仍受限于交互数据的稀缺。

近年来,自监督学习(self-supervised learning, SSL)因能从无标注数据中挖掘监督信号,为缓解数据稀疏性提供了新思路。SGL(self-supervised graph learning)^[6]与 HCCF(hypergraph contrastive collaborative filtering)^[7]聚焦于交互增强,未充分利用多模态特征。为此,部分工作尝试将 SSL 与多模态结合,MMSSL(multimodal self-supervised learning for recommendation)^[8]利用交互信息增强模态表示。然而,现有方法多依赖随机增强和启发式对齐,易引入语义不一致的伪相关信号,影响推荐质量。

为解决上述问题,本文提出了一种新型多模态图对比学习扩散框架。鉴于用户-物品交互数据天然的图结构,首先利用图卷积神经网络(graph convolutional network, GCN)的邻域聚合机制,有效捕获交互图中蕴含的高阶协同信号,从而学习到超越个体交互的、更具上下文感知的用户与物品表示。在此基础上,引入扩散机制,通过正向加噪与逆向去噪构造高质量对比视图,降低噪声干扰。同时,引入用户-用户同构图和项目-项目同构图,缓解稀疏性并挖掘深层语义关联。最终,模型通过自适应融合策略整合强化后的模态特征,并采用融合 BPR(Bayesian personalized ranking)^[9]、去噪和对比损失的联合优化目标,学习泛化性强的用户与物品表示,显著提升推荐性能。

1 图对比学习与推荐扩散相关工作

1.1 图神经网络推荐

图神经网络(graph neural network, GNN)凭借其在建模高阶连通性方面的卓越能力^[10],已迅速成为近年来推荐系统领域的研究热点。例如,NGCF(neural graph collaborative filtering)^[11]设计 GNN 更好地捕捉顶点交互,让模型可以表达高维特征,显式地将协同过滤信号放入到嵌入过程中。LightGCN(light graph convolution network)^[12]通过移除传统 GCN 中的特征转换和非线性激活,仅保

留邻居传播与聚合,简化了图卷积过程。这种轻量化设计在保留协同信号的同时提升了训练效率和嵌入质量。ApeGNN(adaptive node-wise GNN for recommendation)^[13]开发了一种用于信息聚合的节点式自适应扩散机制,其中每个节点都能够根据局部结构自适应地决定其扩散权重。MGDN(mutual-guided dynamic network for image fusion)^[14]可以泛化 LightGCN^[12],并提供灵活的控制来平衡自身和邻居信息。然而,上述模型在整合与利用文本、图像等多模态信息方面尚显不足,这限制了其在处理交互数据稀疏性问题时的建模能力。

1.2 多模态推荐

为解决传统推荐系统固有的数据稀疏问题,诸多研究工作聚焦于融合多模态信息。早期的工作,如 VBPR^[4]和 Deep Style^[15],将项目 ID 嵌入和视觉特征结合用于矩阵分解^[16]。VECF(variational encoder for collaborative filtering)^[17]使用带有注意力机制的 VGG(visual geometry group)^[18]来获取项目的视觉特征。图神经网络有助于多模态用户和项目表示学习。MMGCN(multimodal graph convolutional network)^[19]旨在利用多模态数据和用户-物品之间的交互信息,为个性化推荐提供更准确、全面的解决方案。DualGNN^[5]引入了一个特征偏好模块来捕捉多模态动态特征。以往的方法通常依赖用户-项目交互图来隐式地捕捉项目关系,而部分方法则对项目关系进行显式建模。LATTICE(latent structure mining method for multimodal recommendation)^[20]为每种模态分别执行项目-项目结构学习,然后将输出进行组合。FREEDOM(freezing and denoising graph structures for multimodal recommendation)^[21]通过冻结项目-项目图并去噪用户-项目交互图来简化这一过程。然而,这些方法通常依赖于简单的随机增强或直观的交叉视图对齐,可能会引入不相关的噪声信息。

1.3 用于推荐系统的扩散模型

扩散模型^[22-23]通过模拟数据的逐步损坏与恢复过程,能够有效建模用户偏好的动态生成机制,从而提升推荐系统的性能和鲁棒性。例如,PDRec(plug-in diffusion model for recommendation)^[24]捕捉用户偏好的动态演变来提高扩散模型的性能;DreamRec(diffusion model for recommendation)^[25]引入了扩散模型来探索项目空间中项目的潜在联系,并使用用户的顺序行为作为指导来生成最终推荐;DiffRec(diffusion recommender model)^[26]通过在去噪扩散框架下生成高质量的协同信息以提升

个性化水平; DiffKG(diffusion enhanced knowledge graph for recommendation)^[27] 将扩散模型应用于知识图谱, 通过模拟知识图谱的损坏和恢复过程, 增强用户偏好的编码鲁棒性。

受到上述方法的启发, 本文提出了一种以扩散模型为核心的多模态图对比学习框架, 旨在提升复杂多模态推荐任务中的表示能力。

2 网络结构与方法

DEMGR 由扩散图对比学习、对偶同构图与

多任务联合优化 3 部分组成, 总体设计如图 1 所示。该框架旨在通过模块间的协同增效, 提升推荐表示的鲁棒性。具体而言, 通过扩散图对比学习生成高质量对比视图以学习鲁棒的初始模态表示; 对偶同构图在其基础上进一步建模用户间与物品间的同构关系, 以增强表示的全局感知能力; 多任务联合优化将推荐任务与扩散、对比、对齐等辅助任务统一训练, 通过正则化约束引导模型学习更具泛化性的表征。三者机制上互补, 从而有效提升多模态推荐性能。

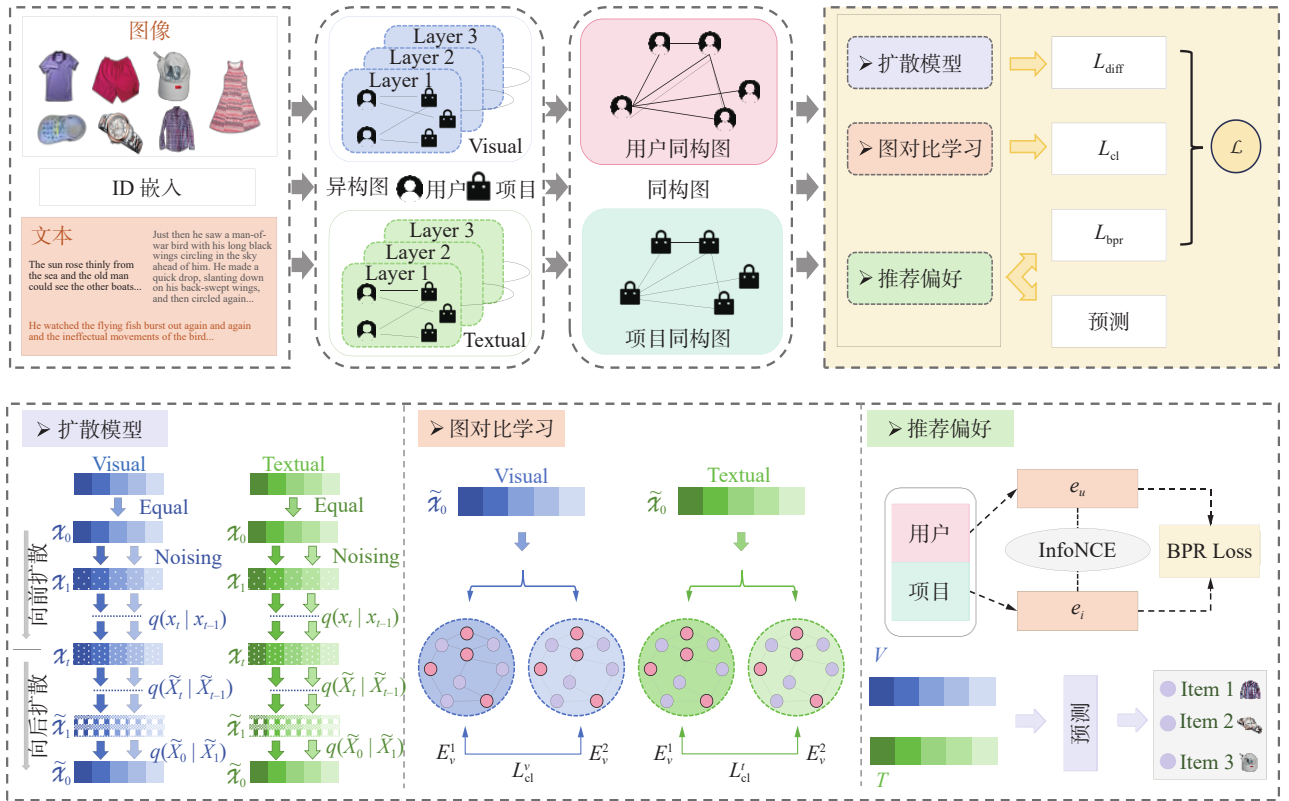


图 1 DEMGR 的总体架构

Fig. 1 Overall architecture of DEMGR

2.1 预备知识和符号定义

首先, 介绍文中涉及的一些基本概念和符号。本文定义 $\mathcal{U} = \{u\}$ 为用户合集, $\mathcal{I} = \{i\}$ 为项目合集。然后, 将每种模态的特征和输入的标识符 (identifier, ID) 嵌入表示为

$$E_m = \{e_m^u \| e_m^i\} \in \mathbf{R}^{d_m \times (|\mathcal{U}| + |\mathcal{I}|)}$$

式中: $m \in \mathcal{M}$ 为模态, $\mathcal{M}$ 为模态的集合, d_m 为特征的维度, $\|$ 为拼接操作。本文仅考虑 ID、视觉和文本这 3 种模态, 记为

$$\mathcal{M} = \{\text{id}, v, t\}$$

2.2 扩散图对比学习

由于扩散模型在高质量数据合成方面具有卓越能力, 一种新颖的基于扩散机制的多模态推荐

框架得以构建。该框架的核心在于一种独特的扩散图对比学习策略。具体而言, 通过在用户-物品交互图中有控制地引入高斯噪声, 模拟了交互信息的扰动过程; 随后, 利用概率扩散模型的逆向过程, 对受扰数据进行去噪处理以重建原始交互状态。此过程不仅旨在恢复信息, 更关键的是能够生成语义相近但存在细微差异的对比视图, 从而有效强化用户与物品的深层特征表示, 提升推荐精度。

2.2.1 交互异构图

为了挖掘不同模态的用户和项目的高阶模态特征, 构建了 3 个用户-物品图 $\mathcal{G} = \{\mathcal{G}_m | \mathcal{G}_v, \mathcal{G}_i, \mathcal{G}_{\text{id}}\}$ 。每个图 $\mathcal{G}_m$ 具有相同的图结构, 并且只保留与每个模态相关的节点特征。其数学表达式为

$$E_m^{(l)} = \sum_{i \in N_u} \frac{1}{\sqrt{|N_u|} \sqrt{|N_i|}} E_m^{(l-1)}$$

式中: N_i 为图 $\mathcal{G}_m$ 中 i 的一跳邻居, N_u 为图 $\mathcal{G}_m$ 中 u 的一跳邻居。

$$E_m = \sum_{l=0}^L E_m^{(l)}$$

式中: L 为图卷积的层数; $E_m^{(0)}$ 为初始特征提取后的原始特征, 信息在图上通过多层传播进行聚合。

2.2.2 向前扩散过程

首先, 捕获的高阶模态特征被定义为 E_m , 它由用户嵌入 e_m^u 和物品嵌入 e_m^i 拼接而成, 其数学形式为

$$E_m = [e_m^u \ e_m^i]$$

前向扩散过程被建模为一个马尔可夫链。在该链中, 通过在每个离散的时间步 t (从 1 到 T) 逐步向数据中添加高斯噪声, 生成一系列加噪后的状态 $x_{1:T}$ 。具体地, 从状态 x_{t-1} 到 x_t 的单步转移概率定义为

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I)$$

式中: $\mathcal{N}$ 为高斯分布, $\beta_t \in (0, 1)$ 为在时间步 t 控制噪声添加量的预设方差调度参数。当时间步 T 足够大时 ($t \rightarrow \infty$), x_T 将近似服从标准高斯分布。

由于高斯噪声的独立可加性, 可以直接从初始状态 x_0 推导出任意时间步 t 的状态 x_t 的分布。

$$q(x_t|x_0) = \mathcal{N}(x_t; \sqrt{\bar{\alpha}_t}x_0, (1-\bar{\alpha}_t)I)$$

进一步地, 任意时刻 t 的加噪状态 x_t 可以通过重参数化技巧表示为

$$x_t = \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon_t$$

为控制在 $x_{0:T}$ 过程中噪声的引入量, 可以采用一个线性的噪声调度器来规划 $1-\bar{\alpha}_t$ (或 $1-\bar{\gamma}_t$) 的值。

$$1-\bar{\alpha}_t = s \cdot \left[\alpha_{\min} + \frac{t-1}{T-1}(\alpha_{\max} - \alpha_{\min}) \right]$$

2.2.3 反向扩散过程

反向过程的目标是学习从噪声 x_T 逐步去除噪声, 恢复出原始数据 x_0 。这一过程使多模态扩散能够有效地捕捉复杂生成过程中的细微变化。反向过程的数学表达式为

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t))$$

式中: $\mu_\theta(x_t, t)$ 为预测的均值, $\Sigma_\theta(x_t, t)$ 为预测的方差。通过训练去噪网络, 模型能够生成与原始数据分布相似的样本。

2.2.4 图对比学习

为了提升表示的质量和鲁棒性, 进一步引入图对比学习。核心思想是利用扩散模型生成的增强视图作为正样本。例如, 对于经过图卷积得到

的视觉模态表示 E_v , 设扩散模型的初始状态 $x_0 = E_v$, 通过前述的扩散和反向去噪过程生成一个语义相近的增强视图 E_v^1 。在对比学习阶段, 采用 InfoNCE 损失^[28] 作为优化目标。对于用户嵌入和物品嵌入分别计算损失。

视觉模态下的用户对比损失 $\mathcal{L}_u^v$ 和物品对比损失 $\mathcal{L}_i^v$ 定义为

$$\mathcal{L}_u^v = \sum_{u_1 \in \mathcal{U}} -\log \frac{\exp(s(e_{u_1,v}^1, e_{u_1,v}^2)/\tau)}{\sum_{u_2 \in \mathcal{U}} \exp(s(e_{u_2,v}^1, e_{u_2,v}^2)/\tau)}$$

$$\mathcal{L}_i^v = \sum_{i_1 \in \mathcal{I}} -\log \frac{\exp(s(e_{i_1,v}^1, e_{i_1,v}^2)/\tau)}{\sum_{i_2 \in \mathcal{I}} \exp(s(e_{i_2,v}^1, e_{i_2,v}^2)/\tau)}$$

视觉模态的总对比损失 $\mathcal{L}_{cl}^v$ 是用户损失和物品损失之和。

$$\mathcal{L}_{cl}^v = \mathcal{L}_u^v + \mathcal{L}_i^v$$

式中: $s(\cdot)$ 为余弦相似度函数, τ 为一个控制模型收敛速率的超参数。同理可得文本模态对比损失。

2.3 对偶同构图

2.3.1 用户同构图

为探究用户间的潜在关联, 构建了用户-用户图。鉴于计算开销, 仅针对具有高度相似偏好的用户子集进行分析。基于此, 所有用户之间的两两余弦相似度计算为

$$S_{u,\hat{u}} = \frac{(e_u)^T \hat{e}_{\hat{u}}}{\|e_u\| \|\hat{e}_{\hat{u}}\|}$$

对于每个用户 u , 保留所有用户对中与 Top-K 个用户相连的边 $S_u^k = \{s_{u,\hat{u}} | u, \hat{u} \in \mathcal{U}\}$ 。

$$s_{u,\hat{u}} = \begin{cases} S_{u,\hat{u}}, & S_{u,\hat{u}} \in \text{top-k}\{S_{u,v} | v \in \mathcal{U}\} \\ 0, & \text{其他} \end{cases}$$

式中: $s_{u,v} | v \in \mathcal{U}$ 为用户 u 对邻居用户 v 的评分。对于构建的用户-用户图 $\mathcal{G}^u = \{\mathcal{U}, S_u^k\}$, 采用 Softmax 函数进行信息传播。

$$A_u^{(l)} = A_u^{(l-1)} + \sum_{\substack{s_{u,\hat{u}} \in S_u^k \\ \hat{u} \in \mathcal{U}}} \frac{\exp(s_{u,\hat{u}})}{\sum_{\substack{s_{u,\hat{u}} \in S_u^k \\ \hat{u} \in \mathcal{U}}} \exp(s_{u,\hat{u}})} A_{\hat{u}}^{(l-1)}$$

式中: $A_u^{(l)}$ 为学习到的用户 u 的第 l 层表示。用户 u 的初始表示向量 $A_u^{(0)}$ 是预处理的用户特征矩阵中第 u 行的行向量。

2.3.2 项目同构图

为提取物品间的显著语义关系采用 KNN 方法, 基于每个模态 $m \in \mathcal{M}$ 的物品特征 e_m^i 建立对应的物品-物品图 $\mathcal{G}_m^i$ 。基于此, 在模态 m 下, 所有物品间的成对余弦相似度计算方式阐述为

$$S_{i,\hat{i}}^m = \frac{(e_m^i)^T \hat{e}_m^{\hat{i}}}{\|e_m^i\| \|\hat{e}_m^{\hat{i}}\|}$$

对于每个物品 i 和模态 m , 保留与物品 i 在模态下 Top- K 个最相似物品所对应的边权重 $s_{i,i}^m$ 。

$$s_{i,i}^m = \begin{cases} S_{i,i}^m, & S_{i,i}^m \in \text{top-}k\{S_{i,j}^m | j \in \mathcal{I}\} \\ 0, & \text{其他} \end{cases}$$

由此构建模态 m 下的物品-物品邻接关系。然后, 在每个模态的物品-物品图 $\mathcal{G}_m^i$ 上聚合多层邻居信息来捕捉该模态下的语义关系。

$$A_{i,m}^{(l)} = \sum_{s_{i,i}^m \in S_{i,m}^k} s_{i,i}^m \cdot A_{i,m}^{(l-1)}$$

式中: $A_{i,m}^{(l)}$ 为从模态 m 的物品语义图 $\mathcal{G}_m^i$ 中学习到的物品 i 的第 l 层表示。物品 i 在模态 m 下的初始表示 $A_{i,m}^{(0)}$ 可以设为 e_m^i 。

2.4 多任务联合优化

2.4.1 推荐偏好学习损失

为了更准确地预测用户个性化偏好, 采用广泛使用的 BPR 损失作为主要推荐任务目标函数。BPR 损失旨在优化模型, 使模型对用户交互过的正样本项目的预测偏好得分显著高于未交互的负样本项目。其数学形式为

$$\mathcal{L}_{\text{bpr}} = \sum_{(u,p,n) \in \mathcal{D}} -\log \sigma(\hat{y}_{u,p} - \hat{y}_{u,n})$$

式中 $\mathcal{D}$ 为训练数据集中所有由用户 u 、其对应的正样本物品 p (用户交互过的项目) 以及负样本物品 n (用户未交互或采样得到的项目) 构成的三元组 (u, p, n) 的集合。

2.4.2 扩散驱动的表达增强损失

为了进一步提升各模态特征表示的质量、鲁棒性与区分度, 模型训练中还整合了基于扩散过程的正则化与增强机制, 这部分主要包含去噪损失和图对比损失两个方面。

去噪损失 ($\mathcal{L}_{\text{diff}}$) 源于扩散模型的逆向去噪过程, 其核心目标是训练一个去噪网络 $\epsilon_\theta(x, t)$, 使其能够准确预测在任意扩散时间步 t 从原始数据 x_0 生成含噪数据 x_t 时所实际添加的噪声 ϵ 。该损失通过计算预测噪声与真实噪声间的均方误差 (mean squared error, MSE) 来量化。

$$\mathcal{L}_{\text{diff}} = \mathbb{E}_{t, x_0, \epsilon} \|\epsilon - \epsilon_\theta(\sqrt{\bar{\alpha}_t}x_0 + \sqrt{1 - \bar{\alpha}_t}\epsilon, t)\|^2$$

2.4.3 多模态语义对齐

在多模态推荐系统中, 不同模态的特征分布通常不一致, 融合过程中往往保留了大量的噪声信息。此外, 一些现有的模态语义对齐方法会破坏历史交互信息, 从而对最终的预测产生不利影响。因此, 提出了一种以稳定的 ID 特征为指导的跨模态对齐方法, 有效利用 ID 嵌入更好地对齐不同模态的语义信息, 确保来自不同模态的信息之间的语义一致性。本文使用高斯分布参数化各模

态特征, 分别计算 ID 模态与视觉模态之间的距离, 并将文本模态特征分布作为损失。公式为

$$\begin{cases} E_{\text{id}} \sim \mathcal{N}(\mu_{\text{id}}, \sigma_{\text{id}}^2) \\ E_v \sim \mathcal{N}(\mu_v, \sigma_v^2) \\ E_t \sim \mathcal{N}(\mu_t, \sigma_t^2) \end{cases}$$

$$\begin{cases} \mathcal{L}_{\text{aligned}_1} = |\mu_{\text{id}} - \mu_v| + |\sigma_{\text{id}} - \sigma_v| \\ \mathcal{L}_{\text{aligned}_2} = |\mu_{\text{id}} - \mu_t| + |\sigma_{\text{id}} - \sigma_t| \end{cases}$$

最终的对齐损失计算公式为

$$\mathcal{L}_{\text{aligned}} = \lambda_{\text{aligned}}(\mathcal{L}_{\text{aligned}_1} + \mathcal{L}_{\text{aligned}_2})$$

式中 λ_{aligned} 为用于平衡对齐损失的超参数。

2.4.4 联合优化目标

图对比学习总损失 $\mathcal{L}_{\text{cl}}$ 计算为

$$\mathcal{L}_{\text{cl}} = \lambda_{\text{cl}}(\mathcal{L}_{\text{cl}}^v + \mathcal{L}_{\text{cl}}^t)$$

式中 λ_{cl} 是用于控制图对比学习损失权重的超参数。

最终, 模型的整体训练是通过最小化多损失联合函数 $\mathcal{L}$ 来驱动的。该总损失函数是推荐偏好学习损失、扩散模型去噪损失及图对比损失的加权和

$$\mathcal{L} = \mathcal{L}_{\text{bpr}} + \alpha \cdot \mathcal{L}_{\text{diff}} + \mathcal{L}_{\text{cl}} + \mathcal{L}_E + \lambda_{\text{aligned}}$$

式中 λ_E 为正则化损失。公式为

$$\mathcal{L}_E = \lambda_E(\|E_v\|_2^2 + \|E_t\|_2^2)$$

式中 α 为一个关键的超参数, 用于平衡核心推荐任务与扩散模型去噪任务之间的相对重要性。值得注意的是, 图对比损失 $\mathcal{L}_{\text{cl}}$ 内部已经包含了其自身的权重因子 λ_{cl} 。通过对这一联合损失函数 $\mathcal{L}$ 进行端到端的优化, 模型得以学习到既能准确预测用户偏好又能生成高质量、信息丰富且具备高泛化能力的多模态特征表示。

3 实验与结果分析

3.1 数据集和评价指标

3.1.1 数据集

实验选用 Baby、Sports 和 Clothing 共 3 个 Amazon 真实数据集, 均经过“5-core”过滤处理。视觉特征为预提取的 4096 维图像向量, 文本特征由预训练的 Sentence Transformer 模型^[29] 提取, 维度为 384。数据统计如表 1 所示。

表 1 实验数据信息
Table 1 Statistics of the datasets

数据集	用户数	物品数	交互数
Baby	19445	7050	160792
Sports	35598	18375	296337
Clothing	39387	23033	278677

3.1.2 评价指标

为公正地评估模型性能,本文采用两种广泛应用的评价指标: $\text{Recall}@K$ 和 $\text{NDCG}@K$, 分别在 $K=10$ 和 $K=20$ 下报告测试数据集中所有用户的平均得分。

3.2 对比方法

为了验证 DEMGR 的性能,与部分基线算法进行比较。这些基线推荐方法可以分为传统单模态推荐和多模态推荐。其中,多模态推荐可以分为仅基于用户-项目图的多模态推荐、基于用户-项目图与项目-项目图的多模态推荐以及基于扩散模型的图神经网络推荐。通过将 DEMGR 与这些算法进行比较,可以证明算法在不同方面的优越性。

3.2.1 传统单模态推荐

1) BPR^[9]: 这是一种针对隐式反馈的推荐法,采用贝叶斯方法建模用户偏好,并通过随机选取负样本提升模型泛化能力。

2) LayerGCN^[12]: 通过动态调整隐藏层表示与自我层的相似性权重,并且采用度敏感的边缘剪枝技术来减少噪声影响。

3.2.2 多模态推荐

1) VBPR^[4]: 该方法是 BPR 模型的一个扩展,引入了物品的视觉特征,旨在提升推荐系统在包含视觉数据的多模态推荐场景下的性能。

2) MMGCN^[19]: 利用 GCN 为每种模态构建特定的图,并通过融合所有模态的结果来计算最终的预测结果。

3) DualGNN^[5]: 该方法使用 GNN 同时对用户和项目之间的关系进行建模,捕获多层次的关系信息,以提高推荐的准确性和个性化。

4) MGCN^[20]: 通过模态感知的对抗扰动和跨模态对比学习捕捉用户偏好。该方法利用 GNN 和对抗性生成网络来增强数据表示,并解决数据稀疏性问题。

5) FREEDOM^[21]: 基于 LATTICE 框架,对用户-项目图去噪并冻项目-项目图,以提升推荐性能。该算法在其基础上创新的加入全局项目高阶语义。

6) MDVT^[30]: 该方法将多模态驱动的虚拟三元组作为额外监督信号,缓解稀疏性并增强模态间对比学习效果。

7) CMDL^[31]: 通过结合对比学习和掩码学习,从用户行为数据中学习表达能力更强的用户及物品表示,从而提升推荐系统的准确性。

8) SMORE^[32]: 通过频域模态融合,抑制模态特有的噪声,提高用户偏好建模的准确性。

9) DiffMM^[33]: 该方法基于扩散模型,通过结合跨模态对比学习以及模态感知的图扩散模型来增强用户表示,从而实现更准确的推荐结果。

3.3 实验环境与参数设置

本文模型采用 Python3.11.5 为开发语言,使用 PyTorch 2.3.0 框架实现。实验所用硬件的操作系统为 Windows 10,显卡为 NVIDIA RTX 4090,运行内存为 120 GB。

为确保评估的公平性,所有模型的用户与项目嵌入维度均统一设置为 64,并采用 Xavier 方法^[34]进行参数初始化。同时,所有模型的优化过程均使用 Adam 优化器。针对本文提出的 DEMGR 模型,设置学习率为 1×10^{-4} ,并且将 GCN 的层数固定为 2。在超参数调优方面,扩散损失权重 λ_{diff} 和 λ_{E} 的取值范围为 $\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$, 对齐损失权重 λ_{aligned} 的取值范围为 $\{0.1, 0.2, 0.3\}$, 温度超参 τ 的取值范围为 $\{0.1, 0.2, 0.3, 0.4, 0.5, 0.6\}$, 丢弃率 dropout 的取值范围为 $\{0.1, 0.2, 0.3, 0.4, 0.5\}$ 。为获得最优结果,对以上超参数进行了网格搜索。

3.4 实验结果与分析

3.4.1 总体实验

本文方法与对比方法的最终的实验结果如表 2 所示。

表 2 DEMGR 与各对比方法性能比较

Table 2 Performance comparison between DEMGR and other comparative methods

评价指标	Baby				Sports				Clothing			
	R@10	R@20	N@10	N@20	R@10	R@20	N@10	N@20	R@10	R@20	N@10	N@20
BPR	0.0382	0.0595	0.0202	0.0263	0.0417	0.0633	0.0232	0.0288	0.0200	0.0295	0.0111	0.0135
LayerGCN	<u>0.0529</u>	<u>0.0820</u>	<u>0.0281</u>	<u>0.0355</u>	<u>0.0594</u>	<u>0.0916</u>	<u>0.0323</u>	<u>0.0406</u>	<u>0.0371</u>	<u>0.0566</u>	<u>0.0200</u>	<u>0.0247</u>
VBPR	0.0423	0.0663	0.0223	0.0284	0.0558	0.0856	0.0307	0.0384	0.0281	0.0415	0.0158	0.0192
MMGCN	0.0378	0.0615	0.0200	0.0261	0.0370	0.0605	0.0193	0.0254	0.0218	0.0345	0.0110	0.0142
DaulGNN	0.0448	0.0716	0.0240	0.0309	0.0568	0.0859	0.0310	0.0385	0.0454	0.0683	0.0241	0.0299
MGCN	0.0616	0.0943	0.0330	0.0414	0.0736	0.1105	0.0403	0.0498	0.0649	0.0971	0.0356	0.0438

续表 2

评价指标	Baby				Sports				Clothing			
	R@10	R@20	N@10	N@20	R@10	R@20	N@10	N@20	R@10	R@20	N@10	N@20
FREEDOM	0.0627	0.0992	0.0330	0.0424	0.0717	0.1098	0.0385	0.0481	0.0629	0.0941	0.0341	0.0420
MDVT	0.0637	0.0990	0.0344	0.0430	0.0736	0.1083	0.0390	0.0495	0.0632	0.0962	0.0347	0.0426
CMDL	0.0643	0.0997	0.0346	0.0434	0.0738	0.1088	0.0394	0.0500	0.0640	0.0956	0.0346	0.0427
SMORE	<u>0.0650</u>	<u>0.0999</u>	<u>0.0348</u>	<u>0.0438</u>	<u>0.0740</u>	<u>0.1109</u>	<u>0.0399</u>	<u>0.0502</u>	<u>0.0650</u>	<u>0.0973</u>	<u>0.0358</u>	<u>0.0440</u>
DiffMM	0.0619	0.0947	0.0326	0.0394	0.0683	0.1019	0.0393	0.0484	—	—	—	—
DEMGR	0.0682	0.1054	0.0363	0.0458	0.0767	0.1148	0.0419	0.0517	0.0667	0.0996	0.0368	0.0450
提升/%	4.92	5.45	4.31	4.57	3.65	3.53	5.01	2.99	2.62	2.36	2.79	2.27

注: 加粗代表最优值, 下划线代表次优值。

在所有评估数据集和评价指标上, DEMGR 均显著优于对比的推荐方法, 包括传统协同过滤模型和先进的多模态推荐模型。在 Baby 数据集上, R@10 提升了 4.92%, R@20 提升了 5.45%, N@10 提升了 4.31%, N@20 提升了 4.57%; 在 Sports 数据集上, R@10 提升了 3.65%, R@20 提升了 3.53%, N@10 提升了 5.01%, N@20 提升了 2.99%; 在 Clothing 数据集上, R@10 提升了 2.62%, R@20 提升了 2.36%, N@10 提升了 2.79%, N@20 提升了 2.27%。

实验结果进一步证实, 基于多模态信息的 GNN 推荐方法优于仅依赖用户-物品交互的通用推荐模型。这一优势在 Clothing 数据集上表现得尤为突出, 在该数据集的 R@10、R@20、N@10 和 N@20 这 4 项评估指标上, 分别取得了高达 81.67%、75.97%、84.00% 和 82.19% 的性能提升。这进一步凸显了视觉、文本等辅助模态在丰富物品表征、缓解稀疏性以及提升用户兴趣建模完整性方面的关键作用。

此外, 集成扩散机制的推荐方法在各项任务中表现更为出色, 验证了高质量增强视图对图对比学习效果的提升作用。扩散模型在捕捉用户潜在偏好、提升表示鲁棒性方面展现出良好潜力。

综上, DEMGR 不仅在性能上取得领先, 其框架设计也验证了扩散机制与多模态融合在图推荐任务中的协同增益效应。

3.4.2 消融实验

为系统性地评估 DEMGR 中各个核心组件的独立贡献, 并验证其设计创新点的有效性与合理性, 设计了一系列消融实验。通过构建并比较 DEMGR 若干简化变体, 旨在量化分析不同组件对整体推荐性能的具体影响。主要的消融变体设置如下:

1) DEMGR_{ui}: 此变体在 DEMGR 完整模型的基础上, 移除了用户-用户同构图和物品-物品同

构图两个模块。

2) DEMGR_i: 在原有算法的基础上, 去掉了项目-项目图。

3) DEMGR_u: 在原有算法的基础上, 去掉了用户-用户图。

4) DEMGR_d: 此变体在扩散模型相关部分, 仅保留了扩散图对比学习任务, 并移除 ID 模态引导的语义对齐任务及所有特征增强任务。

5) DEMGR_{ai}: 此变体侧重于模态对齐机制, 仅保留基于 ID 模态指导的模态内语义对齐任务, 移除扩散图对比学习任务及所有特征增强任务。

消融实验结果如表 3 所示。从表 3 可以发现去掉 DEMGR 的任何模块都会导致结果的下降。

表 3 消融实验
Table 3 Ablation experiment

数据集	方法	R@10	R@20	N@10	N@20
Baby	DEMGR _{ui}	<u>0.0661</u>	<u>0.1028</u>	<u>0.0347</u>	<u>0.0445</u>
	DEMGR _i	0.0546	0.0884	0.0312	0.0388
	DEMGR _{ui}	0.0543	0.0871	0.0298	0.0372
	DEMGR _d	0.0615	0.0957	0.0319	0.0407
	DEMGR _{ai}	0.0647	0.0998	0.0321	0.0435
Sports	DEMGR _{ui}	<u>0.0701</u>	<u>0.1083</u>	<u>0.0380</u>	<u>0.0440</u>
	DEMGR _i	0.0622	0.0760	0.0337	0.0422
	DEMGR _{ui}	0.0613	0.0927	0.0328	0.0415
	DEMGR _d	0.0689	0.1037	0.0377	0.0473
	DEMGR _{ai}	0.0693	0.1067	0.0376	0.0423
Clothing	DEMGR _{ui}	0.0630	0.0945	0.0339	0.0417
	DEMGR _i	0.0493	0.0722	0.0281	0.0330
	DEMGR _{ui}	0.0468	0.0704	0.0264	0.0318
	DEMGR _d	0.0528	0.0821	0.0302	0.0369
	DEMGR _{ai}	<u>0.0648</u>	<u>0.0951</u>	<u>0.0340</u>	<u>0.0419</u>

注: 下划线代表最优值。

首先, $DEMGR_u$ 、 $DEMGR_i$ 以及两者结合的变体 $DEMGR_{ui}$ 均导致了显著的性能下降。其中, $DEMGR_{ui}$ 变体在 Baby 数据集上的 $Recall@10$ 下降最为明显, 相较于 $DEMGR$ 下降了 25.6%。这说明用户间的群体偏好建模与物品间的深层语义关联挖掘对于提升推荐性能至关重要。

其次, 在两种表示增强机制的变体中, $DEMGR_d$ 相较于完整模型有明显差距, 在 Baby 数据集上 $Recall@20$ 指标低了 10.1%, 这说明被移除的 ID 模态引导语义对齐任务, 起到补充监督的作用。同时, 在部分数据集上, 仅保留语义对齐的变体 $DEMGR_{al}$ 性能优于 $DEMGR_d$, 这揭示了两种机制的互补性。

最后, 综合所有消融实验结果来看, 任意核心组件的移除都会导致整体性能的明显折损, 这揭示了模型各模块间存在着协同增益效应。

3.4.3 可视化分析

为直观评估扩散处理机制在多模态表示学习中的有效性, 对 Baby 数据集的用户和物品嵌入进行了 t -SNE^[35] 降维可视化, 结果如图 2 所示。图 2 中对比了模型在应用扩散图对比学习策略前后的多模态表示分布, 其中浅青色、浅蓝色和浅绿色分别代表文本、视觉和 ID 模态。从图 2(a) 中可以清晰地观察到, 未经扩散处理的基线模态表示分布较为分散, 且不同模态的点簇间存在显著的重叠区域, 这表明各模态的语义空间存在混淆。在图 2(b) 中, 经扩散策略增强后的各个模态嵌入点簇更紧凑和内聚, 同时不同模态之间的边界也更为清晰, 分离度显著提高。这证明了扩散图对比学习方法能够成功地促进跨模态表示的对齐, 并学习到更具区分性的高质量特征。

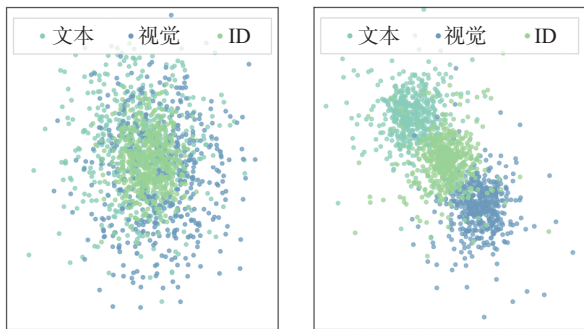


图 2 扩散处理对多模态表示分布的可视化

Fig. 2 Visualizing the distribution of multimodal representations using diffusion processing

3.4.4 实例分析

为直观展示 DEMGR 的优越性, 从 Cloth-

ing 数据集中进行了一项关于用户风格推荐的实例分析。图 3 为 DEMGR 与基线模型 VBPR 的推荐实例对比。

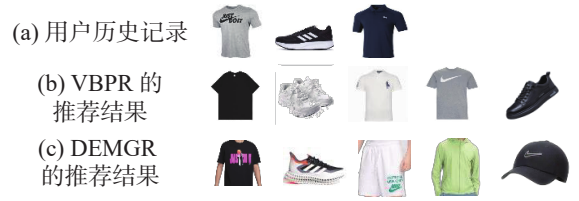


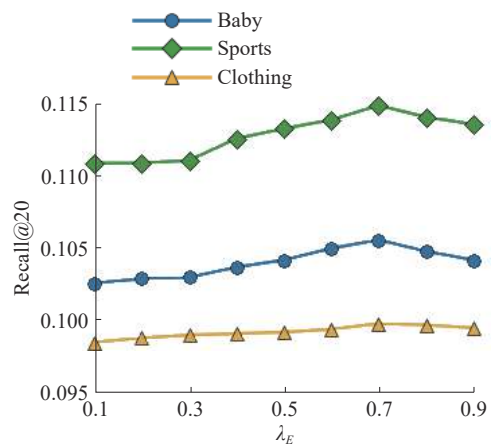
图 3 DEMGR 与基线模型 VBPR 的推荐实例对比
Fig. 3 Comparison of recommended instances between DEMGR and VBPR

图 3(a) 中可以观察到, 该用户的历史偏好体现出运动风格的明确倾向。基线模型 VBPR 的推荐结果如图 3(b), 虽然在品类上具有相关性, 但推荐较为保守, 主要局限于用户已交互过的上衣和鞋子品类中的热门款式, 未能有效拓展用户的兴趣边界。相比之下, 图 3(c) 中 DEMGR 的推荐结果不仅更精准, 且更具多样性。除了推荐新款的同类运动商品外, DEMGR 还成功地跨品类推荐了风格高度一致的运动短裤与配件帽子, 精准捕捉用户对完整运动穿搭的潜在需求。该实例直观地证明了 DEMGR 在挖掘用户潜在兴趣方面的有效性, 并且有效提升复杂多模态推荐任务中的表示能力。

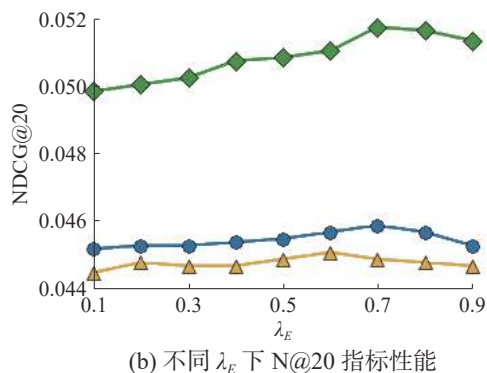
该实例直观地证明了 DEMGR 在挖掘用户潜在兴趣方面的有效性, 并且有效提升复杂多模态推荐任务中的表示能力。

3.4.5 超参数分析

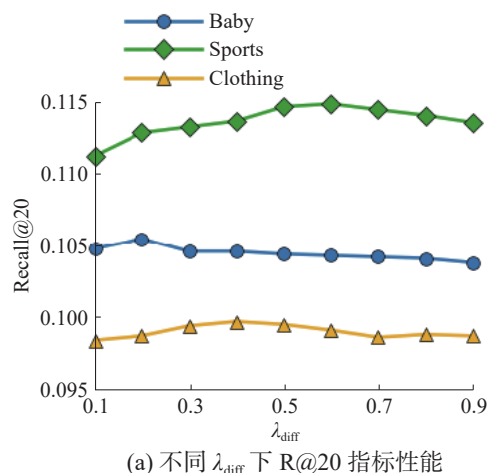
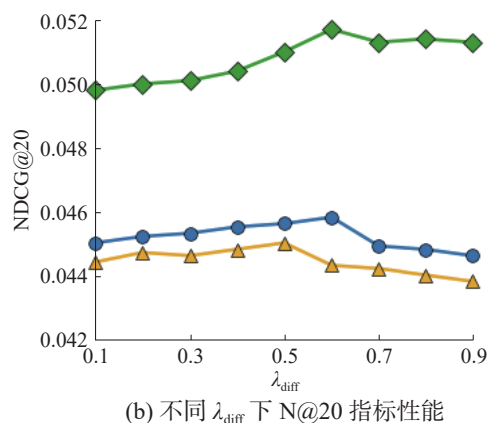
为评估扩散去噪任务的作用, 对超参数 λ_E 和 λ_{diff} 进行了敏感性分析, 针对 λ_E 超参数, 其搜索范围内进行了评估, 结果如图 4 所示。实验表明, 该参数对模型在 3 个数据集上的 $R@20$ 和 $N@20$ 表现均有显著影响。整体趋势表现为性能随参数上升而提升, 在 λ_E 为 0.7 附近达到峰值, 随后略有回落。



(a) 不同 λ_E 下 $R@20$ 指标性能

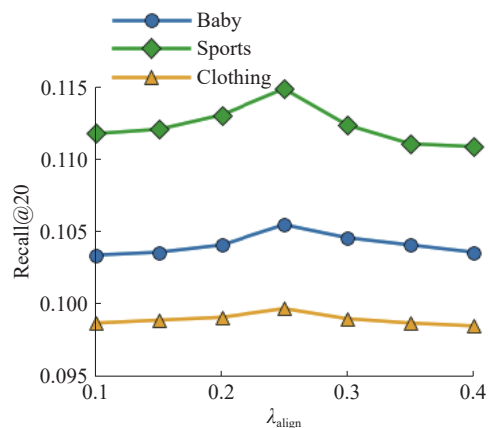
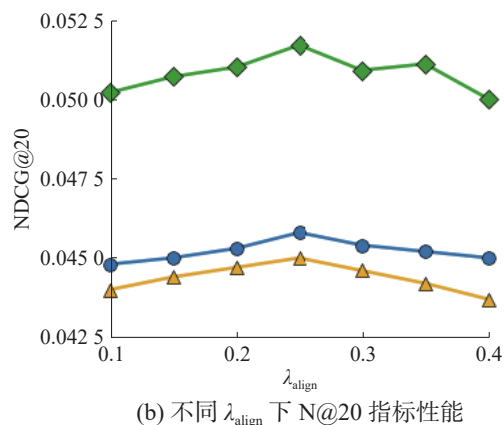
(b) 不同 λ_E 下 N@20 指标性能图 4 超参数 λ_E 的效果Fig. 4 Effect of hyperparameters λ_E

针对超参数 λ_{diff} 分析, 结果如图 5 所示。实验表明, 该参数对模型在 3 个数据集上的 R@20 和 N@20 表现均有显著影响, 整体趋势呈先升后降。性能通常在 λ_{diff} 为 0.5 或 0.6 时达到峰值, 验证了扩散去噪作为正则项在增强表示学习中的有效性。

(a) 不同 λ_{diff} 下 R@20 指标性能(b) 不同 λ_{diff} 下 N@20 指标性能图 5 超参数 λ_{diff} 的效果Fig. 5 Effect of hyperparameters λ_{diff}

分析 λ_{aligned} 在其搜索范围 $\{0.1, 0.2, 0.3\}$ 内变化对模型性能的影响。图 6(a) 和图 6(b) 给出 DEMGR 模型在不同设置下 λ_{aligned} 的性能趋势。实验结果显示, 在 Baby 和 Sport 数据集中, λ_{aligned} 设置为 0.2 时, 模型表现最优; 在 Clothing 数据集中, λ_{aligned}

设置为 0.1 时, 模型表现最优。这突显了模态间对齐在促进多模态信息有效融合中的关键作用。

(a) 不同 λ_{align} 下 R@20 指标性能(b) 不同 λ_{align} 下 N@20 指标性能图 6 超参数 λ_{aligned} 的效果Fig. 6 Effect of hyperparameters λ_{aligned}

4 结束语

本文提出了一种面向多模态推荐任务的图对比学习框架 DEMGR。该框架针对当前多模态推荐中普遍存在的交互数据稀疏与模态噪声问题, 设计了扩散机制辅助的对比视图生成方法, 从源头上提升了自监督学习阶段表示学习的鲁棒性。同时, DEMGR 构建融合用户与物品的同构图结构, 以增强图中节点之间的连接性, 有效缓解数据稀疏问题。在多任务联合优化框架下, 模型能够同时捕捉模态内部的潜在语义信息与模态间的互补特征, 学习出更为精确和稳定的用户与物品表示。大量实验结果表明, DEMGR 在 3 个真实世界数据集上的 R@20 和 N@20 等指标均显著优于当前主流推荐模型, 充分验证了其有效性。

未来, 考虑进一步探索如何利用大语言模型强大的生成能力与世界知识, 来引导扩散过程生成更具可解释性的视图, 并对多模态特征进行更深层次的语义增强, 以推动多模态推荐系统迈向新的性能高度。

参考文献:

- [1] 姜信景, 齐小刚, 刘立芳. 个性化信息推荐方法研究[J]. 智能系统学报, 2018, 13(2): 189195.
JIANG Xinjing, QI Xiaogang, LIU Lifang. Research on the recommendation method of personalized information[J]. CAAI transactions on intelligent systems, 2018, 13(2): 189195.
- [2] 束玮, 李翔, 孙纪舟, 等. 基于动态兴趣传播和知识图谱的推荐方法[J]. 智能系统学报, 2024, 19(4): 9971006.
SHU Wei, LI Xiang, SUN Jizhou, et al. Recommendation method based on dynamic interest propagation and knowledge graph[J]. CAAI transactions on intelligent systems, 2024, 19(4): 9971006.
- [3] 姚暄, 高君宇, 徐常胜. 基于自监督图对比学习的视频问答方法[J]. 软件学报, 2023, 34(5): 20832100.
YAO Xuan, GAO Junyu, XU Changsheng. Video question answering method based on self-supervised graph contrastive learning[J]. Journal of Software, 2023, 34(5): 20832100.
- [4] HE Ruining, MCAULEY J. VBPR: visual Bayesian personalized ranking from implicit feedback[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Phoenix: AAAI, 2016.
- [5] WANG Qifan, WEI Yinwei, YIN Jianhua, et al. DualGNN: dual graph neural network for multimedia recommendation[J]. IEEE transactions on multimedia, 2023, 25: 1074–1084.
- [6] WU Jiancan, WANG Xiang, FENG Fuli, et al. Self-supervised graph learning for recommendation[C]//Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. Online: ACM, 2021.
- [7] XIA Lianghao, HUANG Chao, XU Yong, et al. Hypergraph contrastive collaborative filtering[C]//Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval. Madrid: ACM, 2022.
- [8] WEI Wei, HUANG Chao, XIA Lianghao, et al. Multi-modal self-supervised learning for recommendation[C]//Proceedings of the ACM Web Conference 2023. Austin: ACM, 2023.
- [9] RENDLE S, FREUDENTHALER C, GANTNER Z, et al. BPR: Bayesian personalized ranking from implicit feedback[EB/OL]. (2012-05-09)[2025-05-22]. <https://arxiv.org/abs/1205.2618>.
- [10] 吴国栋, 查志康, 涂立静, 等. 图神经网络推荐研究进展[J]. 智能系统学报, 2020, 15(1): 1424.
WU Guodong, ZHA Zhikang, TU Lijing, et al. Research advances in graph neural network recommendation[J]. CAAI transactions on intelligent systems, 2020, 15(1): 1424.
- [11] WANG Xiang, HE Xiangnan, WANG Meng, et al. Neural graph collaborative filtering[C]//Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. Paris: ACM, 2019.
- [12] HE Xiangnan, DENG Kuan, WANG Xiang, 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. Xi'an: ACM, 2020.
- [13] ZHANG Dan, ZHU Yifan, DONG Yuxiao, et al. ApeGNN: node-wise adaptive aggregation in GNNs for recommendation[C]//Proceedings of the ACM Web Conference 2023. Austin: ACM, 2023.
- [14] GUAN Yuanshen, XU Ruikang, YAO Mingde, et al. Mutual-guided dynamic network for image fusion[C]//Proceedings of the 31st ACM International Conference on Multimedia. Ottawa: ACM, 2023.
- [15] LIU Qiang, WU Shu, WANG Liang. DeepStyle: learning user preferences for visual recommendation[C]//Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. Shinjuku Tokyo: ACM, 2017.
- [16] 田震, 潘腊梅, 尹朴, 等. 深度矩阵分解推荐算法[J]. 软件学报, 2021, 32(12): 39173928.
TIAN Zhen, PAN Lamei, YIN Pu, et al. Deep matrix factorization recommendation algorithm[J]. Journal of software, 2021, 32(12): 39173928.
- [17] CHEN Xu, CHEN Hanxiong, XU Hongteng, et al. Personalized fashion recommendation with visual explanations based on multimodal attention network: towards visually explainable recommendation[C]//Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. Paris: ACM, 2019.
- [18] 范航舟, 梅红岩, 赵勤, 等. 融合双注意力机制的 GNN 多维时间序列预测[J]. 智能系统学报, 2024, 19(5): 12771286.
FAN Hangzhou, MEI Hongyan, ZHAO Qin, et al. Multivariate time series forecasting with a graph neural network and dual attention mechanism[J]. CAAI transactions on intelligent systems, 2024, 19(5): 12771286.
- [19] WEI Yinwei, WANG Xiang, NIE Liqiang, et al. MMGCN: multi-modal graph convolution network for personalized recommendation of micro-video[C]//Proceedings of the 27th ACM International Conference on

- Multimedia. Nice: ACM, 2019.
- [20] ZHANG Jinghao, ZHU Yanqiao, LIU Qiang, et al. Mining latent structures for multimedia recommendation [C]//Proceedings of the 29th ACM International Conference on Multimedia. Online: ACM, 2021.
- [21] ZHOU Xin, SHEN Zhiqi. A tale of two graphs: freezing and denoising graph structures for multimodal recommendation[C]//Proceedings of the 31st ACM International Conference on Multimedia. Ottawa: ACM, 2023.
- [22] 管凤旭, 张涵宇, 路斯棋, 等. 扩散模型在计算机视觉领域的研究现状[J]. 智能系统学报, 2025, 20(2): 265282.
- GUAN Fengxu, ZHANG Hanyu, LU Siqi, et al. Research status of diffusion models in computer vision[J]. *CAAI transactions on intelligent systems*, 2025, 20(2): 265282.
- [23] 蒋世杰, 夏秀山, 翟伟, 等. 基于 ODE 扩散模型的多类异常检测和定位[J]. 智能系统学报, 2025, 20(2): 376388.
- JIANG Shijie, XIA Xiushan, ZHAI Wei, et al. ODE diffusion model for multiclass anomaly detection and localization[J]. *CAAI transactions on intelligent systems*, 2025, 20(2): 376388.
- [24] MA Haokai, XIE Ruobing, MENG Lei, et al. Plug-In diffusion model for sequential recommendation[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Vancouver: AAAI, 2024.
- [25] YANG Zhengyi, WU Jiancan, WANG Zhicai, et al. Generate what you prefer: reshaping sequential recommendation via guided diffusion[C]//Advances in Neural Information Processing Systems 36. New Orleans: Neural Information Processing Systems Foundation, Inc., 2023.
- [26] WANG Wenjie, XU Yiyan, FENG Fuli, et al. Diffusion recommender model[C]//Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval. Taipei: ACM, 2023.
- [27] JIANG Yangqin, YANG Yuhao, XIA Lianghao, et al. DiffKG: knowledge graph diffusion model for recommendation[C]//Proceedings of the 17th ACM International Conference on Web Search and Data Mining. Merida: ACM, 2024.
- [28] YE H C H, HONG Chengyao, HSU Y C, et al. Decoupled contrastive learning[C]//Computer Vision-ECCV 2022. Cham: Springer, 2022.
- [29] REIMERS N, GUREVYCH I. Sentence-BERT: sentence embeddings using siamese bert-networks[EB/OL]. (2019-08-27)[2025-05-22]. <http://arxiv.org/abs/1908.10084>, 2019.
- [30] XU Jinfeng, CHEN Zheyu, LI Jinze, et al. MDVT: enhancing multimodal recommendation with model-agnostic multimodal-driven virtual triplets[C]//Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2. Toronto: ACM, 2025.
- [31] LIN Xixun, LIU Rui, CAO Yanan, et al. Contrastive modality-disentangled learning for multimodal recommendation[J]. *ACM transactions on information systems*, 2025, 43(3): 1-31.
- [32] ONG R K, KHONG A W H. Spectrum-based modality representation fusion graph convolutional network for multimodal recommendation[C]//Proceedings of the Eighteenth ACM International Conference on Web Search and Data Mining. Hannover: ACM, 2025.
- [33] JIANG Yangqin, XIA Lianghao, WEI Wei, et al. DiffMM: multi-modal diffusion model for recommendation[C]//Proceedings of the 32nd ACM International Conference on Multimedia. Melbourne: ACM, 2024.
- [34] KINGMA D P, BA J. Adam: a method for stochastic optimization[EB/OL]. (2014-01-22)[2025-05-22]. <http://arxiv.org/abs/1412.6980>.
- [35] VAN DER MAATEN L, HINTON G. Visualizing data using t-SNE[J]. *Computo a journal of the French Statistical Society*, 2008, 9(11): 1-25.

作者简介:



沈学利, 教授, 博士, 中国计算机学会杰出会员, 软件学院(人工智能学院)院长, 主要研究方向为智能数据处理。获省部级科研成果一等奖 1 项、二等奖 2 项、三等奖 4 项, 获省部级教学成果一等奖 1 项、二等奖 1 项、三等奖 2 项, 发表学术论文近百篇。E-mail: lngd011@163.com。



于佳澜, 硕士研究生, 主要研究方向为推荐系统。E-mail: 1462663058@qq.com。