注意力引导的多步融合：一种用于多模态推荐的层次化融合网络

Yan Zhou† Jie Guo*† Hao Sun†西安电子科技大学 西安电子科技大学 西安电子科技大学 中国陕西西安 中国陕西西安 中国陕西西安 zhouyan123y@stu.xidian.edu.cn jguo@xidian.edu.cn 21011210029@stu.xidian.edu.cn

Bin Song†西安电子科技大学 中国陕西西安 bsong@mail.xidian.edu.cn

Fei Richard Yu‡深圳大学 中国广东深圳 yufei@szu.edu.cn

arXiv:2304.11979v1 [cs.IR] 24 Apr 2023

摘要

多模态推荐的核心思想是合理利用物品的多模态信息来提升推荐性能。以往的工作直接将物品多模态特征与物品ID嵌入相结合，却忽略了多模态特征中包含的内在语义关系。本文提出了一种新颖且有效的多模态推荐注意力引导的多步融合网络，称为 TMFUN。具体而言，我们的模型首先构建模态特征图和物品特征图，以建模潜在的物品间语义结构。然后，我们利用注意力模块识别用户-物品交互数据与多模态数据之间的内在联系，评估多模态数据对不同交互的影响，并实现物品特征的早期融合。此外，我们的模型通过注意力引导的多步融合策略和对比学习优化物品表示，从而提升推荐性能。在三个真实世界数据集上的大量实验表明，与最先进模型相比，我们的模型性能更优。

CCS 概念

• **信息系统** →**推荐系统**。

关键词

特征图，注意力，多步融合，多模态推荐

*通讯作者。†综合业务网国家重点实验室。‡广东省人工智能与数字经济实验室（深圳）。

允许免费复制本作品的全部或部分内容，用于个人或课堂使用，但前提是复制件不得用于营利或商业利益，且必须保留本声明及首页的完整引文。对于本作品中由作者以外的其他方拥有版权的组件，必须尊重其版权。经注明出处后，允许编写摘要。除上述情况外，如需复制、再版、发布到服务器或重新分发至列表，必须事先获得具体许可和/或支付费用。请通过 permissions@acm.org 申请许可。*SIGIR '23, 2023*年*7*月*23*日至*27*日，中国台湾台北© 2023 版权所有归作者/著作权持有人所有。出版权授权予 ACM。ACM ISBN 978-1-4503-9408-6/23/07...$15.00
https://doi.org/10.1145/3539618.3591950

ACM参考文献格式：
Yan Zhou、Jie Guo、Hao Sun、Bin Song 和 Fei Richard Yu。2023。注意力引导的多步融合：用于多模态推荐的层次化融合网络。收录于第*46*届 *ACM* 信息检索研究与发展国际会议论文集（*SIGIR '23*），*2023*年*7*月*23*日至*27*日，中国台湾台北。ACM，纽约，纽约州，美国，5页。https: //doi.org/10.1145/3539618.3591950

1 引言

随着互联网的快速发展，在日常生活中获取多媒体信息变得更加容易，信息过载问题也逐渐演变为严峻挑战。推荐系统（RS）在社交媒体和电子商务领域发挥着重要作用，帮助用户从众多候选项目中挑选出心仪的项目。在推荐系统领域，协同过滤（CF）[8, 19, 22] 是一类传统算法，通过利用用户与项目之间的历史交互来建模用户偏好。协同过滤方法在个性化推荐方面取得了长足进展。此外，NGCF [19] 利用图卷积网络（GCN）对用户-项目历史交互进行编码，并优化用户和物品嵌入。LightGCN [8]改进了GCN的聚合方法，在简化模型的同时提升了推荐性能。

然而，通信和存储系统的发展使项目的信息更加多样化，例如文本、图像、视频等。项目的多模态信息可能会直接影响用户偏好。多模态推荐系统（MRSs）将多模态信息融入传统的协同过滤方法，因应时代需求而产生。早期的多模态推荐 [7, 20] 将多模态数据视为项目的辅助信息，以捕捉用户偏好。近期的研究，例如 [9, 11], 通过注意力机制分析用户的模态偏好。受基于图神经网络（GNN）的推荐方法影响[4, 5, 26], 多模态推荐系统也开始采用GNN来优化多模态特征。Zhang等人 [23] 对视觉和文本信息进行潜在结构学习，以获得项目的多模态潜在结构，然后利用GNN优化多模态表示。Zhang等人 [24]通过GNN传播和聚合多模态信息，然后获得项目的

多模态表示，并通过对比融合实现优化。Mu等人 [14] 提出了一种混合聚类图卷积网络，以捕捉各种用户行为模式，从而提升推荐性能。

尽管基于图神经网络的多模态推荐系统取得了显著成果，但仍存在一些局限性。首先，现有的大多数研究通过用户－物品－用户图挖掘物品间的协同关系，却忽略了物品中丰富的多模态特征。各模态的预训练嵌入反映了模态特定特征，而这些特征无法从交互数据中获取。因此，可以利用预训练视觉和文本特征构建模态特征图和物品特征图，从而获得物品的多模态语义关系。其次，现有的大多数多媒体推荐系统直接将物品多模态数据用作辅助信息，并分析多模态数据对用户偏好的整体影响。事实上，多模态数据对不同交互中用户偏好的影响并非恒定。注意力融合模块 [2] 可用于分析多模态数据在每次交互中的重要程度。交互数据和多模态数据都被视为影响用户偏好的因素。此外，大多数研究没有考虑模态差异。显然，仅通过拼接无法充分挖掘多模态数据中的潜在特征关系，因为视觉和文本数据之间天然存在巨大的语义鸿沟。为了充分利用多模态数据，我们提出了一种层次化融合网络，用于融合不同粒度的多模态特征。此外，我们采用对比学习实现多模态数据的模态间对齐，从而消除视觉和文本数据之间的语义鸿沟。

基于上述分析，本文提出了一种用于多模态推荐的新型注意力引导的多步融合网络，简称为***TMFUN***。具体而言，在所提出的TMFUN模型中，我们首先构建模态特征图和物品特征图，以挖掘丰富的多模态特征。然后，我们设计了一个注意力模块，用于探索交互数据与物品多模态特征之间的潜在关联，分析多模态数据在不同交互中的作用，并实现物品特征的早期融合。此外，我们设计了一种注意力引导的多步融合策略，用于融合不同粒度的物品特征。最后，我们利用对比学习实现多模态数据的模态间对齐。我们的主要贡献可总结为以下三点:

- 我们构建模态特征图和物品特征图，以提取丰富的多模态特征，并设计一个注意力模块，实现交互数据与多模态特征之间的关系挖掘。
- 我们设计了一种注意力引导的多步融合策略，以实现物品特征的多粒度融合，并通过对比学习实现多模态数据的跨模态对齐。
- 所提出的 TMFUN 优于当前最先进的多媒体推荐系统，证明了我们方法的有效性。

2 所提出的方法

本节将详细介绍所提出的模型，如图 1 所示。

2.1 图构建与初始化

2.1.1 用户-物品交互图。 本文将用户-物品交互图定义为 $G_H = \{(u,\ y_{ui},\ i)\ |(u \in U,\ i \in I)\}$，其中 U表示用户集合，I表示物品集合，y_{ui}表示用户 u与物品 i之间的连接关系。当用户 u与物品 i之间存在显式关系时，$y_{ui} = 1$；否则，$y_{ui} = 0$。遵循主流推荐方法 [18, 23],，我们初始化交互图中的用户和物品嵌入，以获得 ID嵌入 $x^u,\ x^i \in R^d$，其中 d为维度。

2.1.2 物品特征图。本文采用由预训练卷积神经网络提取的公开视觉特征 [6],，记为 $x^v \in R^{d_v}$，其中 $d_v = 4096$ 为视觉特征维度。为充分挖掘图像信息之间的关联，我们基于特征相似度构建视觉特征图。按照 [1, 15],，我们使用余弦函数计算物品对 i 和 j之间的相似度分数，

$$s_{ij}^v = \frac{\left(\mathrm{x}_i^v\right)^\top \mathrm{x}_j^v}{\left\|\mathrm{x}_i^v\right\| \left\|\mathrm{x}_j^v\right\|}, \tag{1}$$

其中，x_i^v和x_j^v分别表示物品 i和 j的视觉特征，s_{ij}^v表示物品对 i和 j之间的相似度分数。

然后，对于每个 i，我们仅保留相似度最高的前-N条边的连接，以获得视觉特征图邻接矩阵 S^v。文本特征 $x^t \in R^{d_t}$ $d_t = 378$ 由预训练的 sentence-transformers [16]提取。随后，与视觉特征图类似，我们可以获得文本特征图邻接矩阵S^t。最后，我们使用冻结融合策略，将视觉特征图和文本特征图进行融合，以获得物品特征图邻接矩阵 S。

2.2 注意力关系挖掘

GCN 的消息传播和聚合机制可以优化用户和项目的表示。因此，我们分别在用户-物品交互图、视觉特征图、文本特征图和物品特征图上执行图卷积操作，以获得相应的输出$\tilde{x}^u$、 $\tilde{x}^i$、$\tilde{x}^v$、 $\tilde{x}^t$和 $\tilde{x}^m$。

然后，我们设计了一种模态感知注意力机制，以识别每种模态 q对每次用户-物品交互的贡献。使用 $\tilde{x}^u$ 作为查询，使用 $\tilde{x}^i, \tilde{x}^v, \tilde{x}^t, \tilde{x}^m$ 作为键和值，我们计算每种模态对每次交互的影响 a_{ij}^q，

$$a_{ij}^q = \frac{\exp\left(\frac{\tilde{x}^u \odot \tilde{x}^q}{\sqrt{d}}\right)}{\sum_{q \in Q} \exp\left(\frac{\tilde{x}^u \odot \tilde{x}^q}{\sqrt{d}}\right)}, \tag{2}$$

其中 $q \in Q = \{i,\ v,\ t,\ m\}$ 表示模态。 $\tilde{a}_{ij}^q = \frac{\tilde{x}^u \odot \tilde{x}^q}{\sqrt{d}}$, $\tilde{x}^u$表示用户特征，$\tilde{x}^q$ 是该模态的物品特征 q, $\odot$ d 分别表示点积和转换因子。然后，我们可以得到不同模态之间的交互，并获得融合了其他模态的物品嵌入 $h_{ij}{}^q$，

$$h_{ij}{}^q = \sum_{q \in Q} a_{ij}^q \tilde{x}_j^q. \tag{3}$$

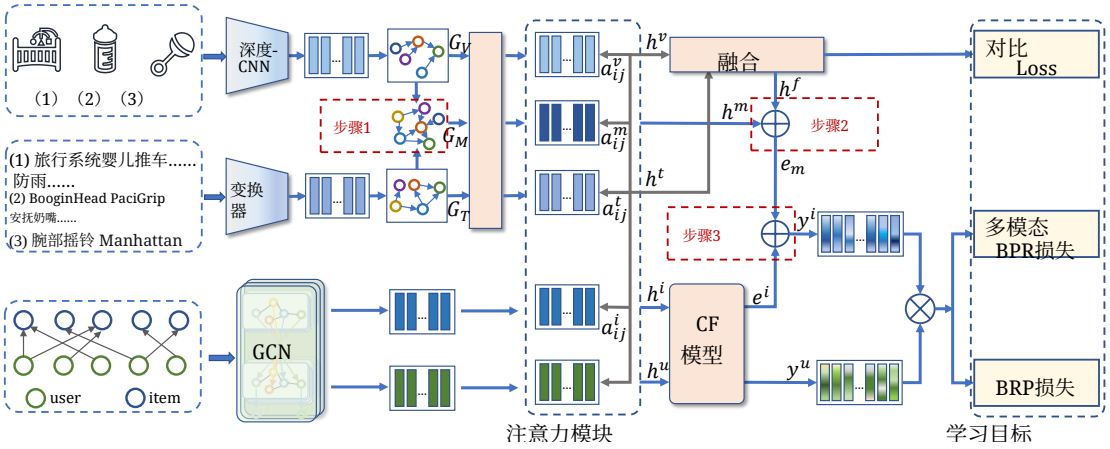

图1：所提出的TMFUN模型示意图。

2.3 多步融合

2.3.1 模态特征图融合。为了降低学习融合系数带来的模型开销，我们采用冻结融合策略 [25]，融合视觉特征图和文本特征图，以获得物品-物品特征图 G_M。

$$G_M = \lambda G_V + (1-\lambda)\, G_T, \tag{4}$$

其中，λ 是图融合系数，G_V 是视觉特征图，G_T 是文本特征图。

2.3.2 跨模态对比融合。类似地，我们融合视觉特征和文本特征，然后使用对比损失来监督视觉模态和文本模态的对齐。随后，我们可以得到物品多模态特征 e^m:

$$e^m = h^m + h^f, \tag{5}$$

其中，h^m 是通过注意力模块获得的物品多模态特征。$h^f = \mu h^v + (1-\mu) h^t$ 是通过对比融合获得的物品融合特征。其中，μ 是对比融合系数，h^v 是视觉特征，h^t 表示文本特征。

2.3.3 物品全局融合。然后，我们融合物品多模态特征 e^m 和从协同过滤模型 [8]中获得的物品特征 e^i，以得到最终物品特征 y^i，

$$y^i = e^i + e^m. \tag{6}$$

2.3.4 偏好预测。为了预测用户的偏好，我们首先根据内积计算用户 u 与项目i 之间的相似度分数，

$$\hat{y}(u,i) = (y^u)^\top y^i, \tag{7}$$

其中 y^u 为用户特征，y^i 为物品特征。然后，根据相似度分数对项目进行排序，并向用户推荐排名前 K 的项目，从而实现 top-K 推荐。

2.4 学习目标

2.4.1 对比损失。为了对齐各模态，并自适应地捕获多个模态间的共享特征，我们设计了对比损失 L_C:

$$\mathcal{L}_C = -\frac{1}{|I|}\sum_{i\in I}\left[\frac{1}{|Q'|}\sum_{q'\in Q'}\frac{1}{2}\left(MI\left(h_i^{q'}, h_i^i\right) + MI\left(h_i^i, h_i^{q'}\right)\right)\right], \tag{8}$$

其中，i表示项目，I表示项目集合，$q_{'}\in Q_{'}=\{v,t\}$表示模态。$MI(\cdot,\cdot)$是互信息函数，用于衡量特征之间的一致性。h_i^i是项目 i 的特征，而 $h_i^{q'}$ 表示项目 i 的 $q_{'}$ 模态特征。

2.4.2 BPR 损失。我们采用 BPR 损失 [17]来优化模型，确保正样本的偏好分数高于负样本。

$$\mathcal{L}_{BPR} = -\sum_{(u,i,j)\in\mathcal{D}} \log\sigma\left(\hat{y}_{ui} - \hat{y}_{uj}\right), \tag{9}$$

其中 $D=\{(u,\ i,\ j)|(u,\ i)\in P,\ (u,\ j)\in N\}$ 是训练集中的三元组实例，P 是正样本集，N 是负样本集，$\sigma(\cdot)$ 表示 Sigmoid 激活函数。

2.4.3 多模态 *BPR* 损失。为了保留模态特定特征，我们设计了一种多模态 BPR 损失，用于优化项目的每种模态特征，

$$\mathcal{L}_{mmbpr} = \sum_{(u,i,j)\in\mathcal{D}}\sum_{q'\in\{v,t\}} -\log\sigma\left((y^u)^\top h_i^{q'} - (y^u)^\top h_j^{q'}\right), \tag{10}$$

其中 $D=\{(u,\ i,\ j)|(u,\ i)\in P,\ (u,\ j)\in N\}$ 是训练集中的三元组实例。$h_i^{q'}$ 是模态 $q_{'}\in\{v,\ t\}$ 的多模态特征，其中 v表示视觉模态，t表示文本模态。$\sigma(\cdot)$ 是 Sigmoid 激活函数。最后，我们可以得到学习目标 L，

$$\mathcal{L} = \mathcal{L}_{BPR} + \alpha\mathcal{L}_{mmBPR} + \beta\mathcal{L}_C, \tag{11}$$

其中，α 是多模态 BPR 损失 L_{mmBPR}的权重系数，β 是对比损失 L_C的权重系数。

表 1：我们的 TMFUN 与其他模型在 Recall@20 (R@20) 和 NDCG@20 (N@20) 指标上的性能比较。最佳性能以粗体突出显示。

方法	Baby		运动		服装	
	R@20	N@20	R@20	N@20	R@20	N@20
矩阵分解[22]	0.0440	0.0200	0.0430	0.0202	0.0191	0.0088
NGCF[19]	0.0591	0.0261	0.0695	0.0318	0.0387	0.0168
LightGCN[8]	0.0698	0.0319	0.0782	0.0369	0.0470	0.0215
IMP-GCN[12]	0.0720	0.0319	0.0806	0.0378	0.0466	0.0211
VBPR[7]	0.0486	0.0213	0.0582	0.0265	0.0481	0.0205
MMGCN[21]	0.0640	0.0284	0.0638	0.0279	0.0507	0.0221
GRCN[20]	0.0754	0.0336	0.0833	0.0377	0.0631	0.0276
LATTICE[23]	0.0829	0.0368	0.0915	0.0424	0.0710	0.0316
MICRO[24]	0.0865	0.0389	0.0968	0.0445	0.0782	0.0351
HCGCN[14]	0.0922	0.0415	0.1032	0.0478	0.0810	0.0370
TMFUN	**0.1032**	**0.0452**	**0.1112**	**0.0496**	**0.0962**	**0.0431**

3 实验

3.1 实验设置

3.1.1 数据集与参数设置。我们在亚马逊评论数据集[13]中三个广泛使用的类别子集上进行实验：(a) 婴儿，(b) 运动与户外，以及(c) 服装、鞋履和珠宝。为简便起见，我们分别用婴儿、运动和服装表示它们。

遵循主流方法 [14, 23], ，我们采用广泛使用的召回率@K和归一化折损累计增益@K作为评估指标，分别简称为R@K和N@K。具体而言，我们展示K=20时的性能。对于每位用户，我们随机选取80%的历史交互用于训练，10%用于验证，剩余10%用于测试。我们的模型采用PyTorch实现，并在RTX3090 GPU显卡上进行评估。我们采用Xavier方法[3] 初始化模型参数，并使用Adam [10] 作为优化器。此外，在我们的模型中，超参数 λ 和 μ 的取值范围为0到1， α 和 β 则从 [1e-4、1e-3]中选取。此外，为确保模型收敛，早停轮数和总训练轮数分别设置为20和1000。

3.2 实验结果分析

3.2.1 性能比较。 性能比较结果如表2所示。总体而言，我们提出的TMFUN模型优于所有其他模型。具体而言，与最先进的模型HCGCN相比，TMFUN模型在婴儿、运动和服装数据集上的R@20分别提升了11.93%、7.75%和18.77%。与现有模型相比，我们的模型通过构建模态特征图，进一步挖掘了丰富的多模态特征

以及物品特征图，这显著提升了多模态推荐的性能。

3.2.2 消融实验。本节如表2所示，在三个数据集上进行了消融实验。我们分别分析了协同过滤方法、注意力模块和多模态特征的影响。TMFUN/MF、TMFUN/NGCF 和 TMFUN 分别表示 TMFUN模型中采用的不同协同过滤方法（MF、NGCF、LightGCN）。TMFUN-mm 和 TMFUN-att 分别表示禁用多模态特征和注意力模块。

表2：我们的TMFUN模型及其变体的性能。

方法	Baby		运动		服装	
	R@20	N@20	R@20	N@20	R@20	N@20
TMFUN/MF	0.0852	0.0379	0.0888	0.0381	0.0828	0.0366
TMFUN/NGCF	0.0812	0.0345	0.0924	0.0397	0.0863	0.0365
TMFUN-mm	0.0828	0.0370	0.0913	0.0406	0.0568	0.0252
TMFUN-att	0.1022	0.0442	0.1106	0.0489	0.0930	0.0417
TMFUN	0.1032	0.0452	0.1112	0.0496	0.0962	0.0431

我们可以观察到，选择适当的协同过滤方法会对模型性能产生显著影响。在 TMFUN模型中，我们采用当前效果较好且轻量的 LightGCN 作为协同过滤模型。此外，多模态特征的利用以及注意力模块的引导都对模型性能产生了积极影响。

3.2.3 参数分析。本节如图 2 所示，对学习目标 L 的参数 α 和 β 进行实验验证与分析，以评估它们的影响。 α 和 β 的参数取值范围为 [1e-5、1e-4、1e-3、1e-2]。从图 2 可以看出，对于 Baby 数据集，当 α 和 β 均为 1e-3 时，模型表现最佳。对于 Clothing 数据集， α 和 β 的最优参数为 1e-4。因此，多模态 BPR 损失和对比损失对模型的成功至关重要，使模型能够充分利用潜在多模态信息。

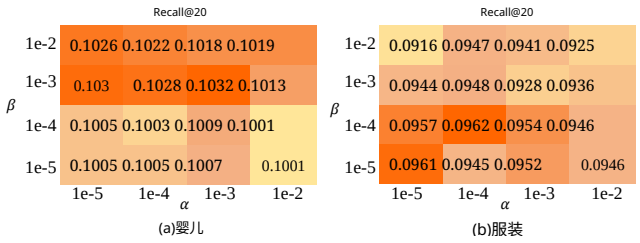

图 2：TMFUN 在 Baby 和 Clothing 数据集上，参数 α 和 β 取不同值时的 Recall@20 表现。

4 结论

为充分利用多模态特征中蕴含的固有语义关系，我们提出了一种用于多模态推荐的新型网络，称为 TMFUN。TMFUN模型的核心思想是通过多模态数据增强物品表示，并为下游协同过滤模型提供更有效的偏好挖掘基础。在特征层面，特征图的构建可以保留模态特定特征，而对比融合可以捕获跨模态共享特征。在架构层面，我们设计了一种注意力引导的多步融合策略，以捕获交互数据与多模态数据之间的潜在关系。在三个真实世界数据集上的大量实验表明，我们提出的 TMFUN模型具有有效性。为了进一步挖掘潜在

的项目间关系，我们将在未来的工作中引入更多模态，例如视频和音频。

致谢

本研究得到中国国家自然科学基金（项目编号：62201424和62271324）、陕西省重点研发计划（项目编号：2023-YBGY-218）的资助，同时得到综合业务网理论及关键技术国家重点实验室和西安电子科技大学高性能计算平台的支持。

参考文献

[1] Jie Chen、Haw-ren Fang 和 Yousef Saad。2009 年。通过递归 Lanczos 二分法快速近似构建高维数据的 K 近邻图。机器学习研究期刊 10, 9（2009）。[2] Yimian Dai、Fabian Gieseke、Stefan Oehmcke、Yiquan Wu 和 Kobus Barnard。2021 年。注意力特征融合。载于*IEEE/CVF* 冬季计算机视觉应用会议论文集（*WACV*）。3560–3569。[3] Xavier Glorot 和 Yoshua Bengio。2010 年。理解深度前馈神经网络训练的困难。载于第十三届人工智能与统计国际会议论文集。JMLR 研讨会与会议论文集，249–256。[4] Jie Guo、Yan Zhou、Peng Zhang、Bin Song 和 Chen Chen。2021 年。基于异构多关系图融合的信任感知推荐。信息融合 74（2021），87–95。[5] Wei Guo、Yang Yang、Yaochen Hu、Chuyuan Wang、Huifeng Guo、Yingxue Zhang、Ruiming Tang、Weinan Zhang 和 Xiuqiang He。2021 年。采用混合归一化的深度图卷积网络，实现准确且多样化的推荐。载于*KDD* 高维稀疏数据深度学习实践第三届研讨会论文集。[6] Ruining He 和 Julian McAuley。2016 年。起伏变化：利用单类协同过滤建模时尚趋势的视觉演化。载于第二十五届国际万维网会议论文集。507–517。[7] Ruining He 和 Julian McAuley。2016 年。VBPR：基于隐式反馈的视觉贝叶斯个性化排序。载于美国人工智能学会会议论文集，第 30 卷。[8] Xiangnan He、Kuan Deng、Xiang Wang、Yan Li、Yongdong Zhang 和 Meng Wang。2020 年。LightGCN：简化并增强用于推荐的图卷积网络。载于第 *43* 届国际 *ACM SIGIR* 信息检索研究与发展会议论文集。639–648。[9] Taeri Kim、Yeon-Chang Lee、Kijung Shin 和 Sang-Wook Kim。2022 年。MARIO：用于多媒体推荐的模态感知注意力与模态保持解码器。载于第 *31* 届 *ACM* 国际信息与知识管理会议论文集。993–1002。[10] Diederik P Kingma 和 Jimmy Ba。2014 年。Adam：一种随机优化方法。*arXiv* 预印本 *arXiv:1412.6980*（2014）。[11] Fan Liu、Huilin Chen、Zhiyong Cheng、Anan Liu、Liqiang Nie 和 Mohan Kankanhalli。2022 年。用于推荐的解耦多模态表示学习。*IEEE* 多媒体汇刊 （2022）。[12] Fan Liu、Zhiyong Cheng、Lei Zhu、Zan Gao 和 Liqiang Nie。2021 年。用于推荐的兴趣感知消息传递图卷积网络。载于*2021* 年万维网会议论文集。1296–1305。[13] Julian McAuley、Christopher Targett、Qinfeng Shi 和 Anton Van Den Hengel。2015 年。基于图像的风格与替代品推荐。载于第 *38* 届国际 *ACM SIGIR* 信息检索研究与发展会议论文集。43–52。[14] Zongshen Mu、Yueting Zhuang、Jie Tan、Jun Xiao 和 Siliang Tang。2022 年。学习多媒体推荐中的混合行为模式。载于第 *30* 届 *ACM* 国际多媒体会议论文集。376–384。[15] Leif E Peterson。2009 年。K 近邻。*Scholarpedia* 4, 2（2009），1883。[16] Nils Reimers 和 Iryna Gurevych。2019 年。Sentence-BERT：使用孪生 BERT 网络的句子嵌入。载于*2019* 年自然语言处理经验方法会议与第九届自然语言处理国际联合会议（*EMNLP-IJCNLP*）论文集。3982–3992。[17] Steffen Rendle、Christoph Freudenthaler、Zeno Gantner 和 Lars Schmidt-Thieme。2009 年。BPR：基于隐式反馈的贝叶斯个性化排序。载于第二十五届人工智能不确定性会议论文集。452–461。[18] Xiang Wang、Xiangnan He、Yixin Cao、Meng Liu 和 Tat-Seng Chua。2019 年。KGAT：用于推荐的知识图谱注意力网络。载于第 *25* 届 *ACM SIGKDD* 国际知识发现与数据挖掘会议论文集。950–958。[19] Xiang Wang、Xiangnan He、Meng Wang、Fuli Feng 和 Tat-Seng Chua。2019 年。神经图协同过滤。载于第 *42* 届国际 *ACM SIGIR* 信息检索研究与发展会议论文集。165–174。

[20] Yinwei Wei、Xiang Wang、Liqiang Nie、Xiangnan He 和 Tat-Seng Chua。2020 年。结合隐式反馈的多媒体推荐图细化卷积网络。载于第 *28* 届 *ACM* 国际多媒体会议论文集。3541–3549。[21] Yinwei Wei、Xiang Wang、Liqiang Nie、Xiangnan He、Richang Hong 和 Tat-Seng Chua。2019 年。MMGCN：用于个性化微视频推荐的多模态图卷积网络。载于第 *27* 届 *ACM* 国际多媒体会议论文集。1437–1445。[22] Hong-Jian Xue、Xinyu Dai、Jianbing Zhang、Shujian Huang 和 Jiajun Chen。2017 年。用于推荐系统的深度矩阵分解模型。载于*IJCAI*，第 17 卷。澳大利亚墨尔本，3203–3209。[23] Jinghao Zhang、Yanqiao Zhu、Qiang Liu、Shu Wu、Shuhui Wang 和 Liang Wang。2021 年。挖掘多媒体推荐的潜在结构。载于第 *29* 届 *ACM* 国际多媒体会议论文集。3872–3880。[24] Jinghao Zhang、Yanqiao Zhu、Qiang Liu、Mengqi Zhang、Shu Wu 和 Liang Wang。2022 年。通过对比模态融合挖掘潜在结构。*IEEE* 知识与数据工程汇刊 （2022）。[25] Xin Zhou。2022 年。两个图的故事：面向多模态推荐的图结构冻结与去噪。*arXiv* 预印本 *arXiv:2211.06924*（2022）。[26] Yan Zhou、Jie Guo、Bin Song、Chen Chen、Jianglong Chang 和 Fei Richard Yu。2022 年。用于推荐的信任感知多任务知识图谱。*IEEE* 知识与数据工程汇刊 （2022）。