

面向国潮文化偏好的多模态电商推荐技术综述

田文芳, 李佳燕, 郑丹, 陈静怡, 蒋智贤, 何庆

贵州大学大数据与信息工程学院, 贵州 贵阳

收稿日期: 2026年3月3日; 录用日期: 2026年3月22日; 发布日期: 2026年4月3日

摘要

随着国潮品牌与文化消费的持续发展, 电商平台中的商品推荐面临由功能与价格导向向文化符号、审美风格与情感认同导向转变的趋势。国潮商品通常同时包含视觉风格、文本语义与文化属性等多源信息, 单一模态或传统协同过滤方法难以有效刻画其复杂的偏好特征。近年来, 基于深度学习的多模态推荐技术在电商领域取得了广泛应用, 为融合商品图像、文本描述、用户评论及结构化属性信息提供了新的研究思路, 但其在国潮文化偏好建模中的系统性总结仍相对不足。本文围绕面向国潮文化偏好的多模态电商推荐技术展开综述。首先分析国潮消费场景下用户偏好与商品特征的多模态特性, 梳理相关数据形态与常用公开数据资源; 随后从多模态表征学习、跨模态融合与对齐、图结构建模与知识增强、模型优化与训练策略四个维度, 对近年来的代表性研究工作进行分类总结, 并探讨各方法在国潮场景下的适配性。在此基础上, 本文探讨了大语言模型驱动的生成式推荐范式, 分析其在深层文化知识推理与可信解释生成方面的破局作用。最后, 本文总结了当前研究在文化语义建模、模态对齐、偏好动态建模与推荐解释等方面面临的主要挑战, 并对多模态大模型驱动的电商推荐、人机协同交互机制以及文化偏好可控建模等未来研究方向进行了展望。本文可为国潮商品推荐系统的研究与实践提供参考。

关键词

国潮文化, 多模态推荐, 深度学习, 电商推荐系统

A Review of Multimodal E-Commerce Recommendation Technologies for Domestic Trend Cultural Preferences

Wenfang Tian, Jiayan Li, Dan Zheng, Jingyi Chen, Zhixian Jiang, Qing He

College of Big Data and Information Engineering, Guizhou University, Guiyang Guizhou

Received: March 3, 2026; accepted: March 22, 2026; published: April 3, 2026

文章引用: 田文芳, 李佳燕, 郑丹, 陈静怡, 蒋智贤, 何庆. 面向国潮文化偏好的多模态电商推荐技术综述[J]. 图像与信号处理, 2026, 15(2): 196-211. DOI: 10.12677/jisp.2026.152017

Abstract

With the continued growth of domestic trend brands and cultural consumption, product recommendations on e-commerce platforms are shifting from a focus on functionality and price to an emphasis on cultural symbols, aesthetic styles, and emotional resonance. Domestic trend products typically involve multi-source information, including visual style, textual semantics, and cultural attributes. As a result, single-modality methods and traditional collaborative filtering approaches often struggle to capture such complex preference patterns effectively. In recent years, deep learning-based multimodal recommendation techniques have been widely adopted in e-commerce, providing new avenues for integrating product images, textual descriptions, user reviews, and structured attribute information. However, systematic reviews of their applications to modeling preferences in domestic trend culture remain limited. This paper presents an overview of multimodal recommendation technologies for e-commerce that are tailored to domestic trend cultural preferences. First, we examine the multimodal characteristics of user preferences and product attributes in domestic trend consumption scenarios, and we identify relevant data modalities, formats, and commonly used public datasets. We then categorize and summarize representative studies from recent years across four dimensions: multimodal representation learning, cross-modal fusion and alignment, graph structure modeling and knowledge enhancement, and model optimization with training strategies, while discussing the suitability of these methods for domestic trend scenarios. Furthermore, this paper highlights the LLM-driven generative recommendation paradigm, analyzing its breakthrough role in deep cultural knowledge reasoning and the generation of trustworthy explanations. Finally, we summarize key challenges in cultural semantic modeling, modality alignment, preference dynamics modeling, and recommendation explainability. We also outline promising future directions, such as multimodal large-model-driven e-commerce recommendation, human-machine collaborative interaction mechanisms, and controllable modeling of cultural preferences. This paper aims to serve as a reference for both research and practice in recommendation systems for domestic trend products.

Keywords

Domestic Trend Culture, Multimodal Recommendation, Deep Learning, E-Commerce Recommendation System

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在数字经济与消费升级的双重驱动下，电子商务已成为国潮品牌触达消费者的核心阵地。然而，平台商品数量的指数级增长使得用户陷入“选择困境”，信息过载问题日益突出，个性化推荐系统因此成为缓解信息过载、连接供需两端的核心技术引擎。但随着国潮消费的兴起，用户决策逻辑已从功能价格导向转向对商品所承载的文化符号、美学价值及身份认同的综合追求，这无疑对传统推荐范式提出了全新挑战[1]。

国潮消费的核心在于用户对商品文化符号的解码与认同。这些符号在视觉层面体现为传统纹样与色彩体系，在叙事层面融入历史典故与工艺传承，在价值层面传递匠心精神与民族自信。相关研究表明，文化符号与情感认同在消费决策中对用户偏好具有显著影响[2]。以“敦煌联名系列”为例，其广受欢迎

既源于藻井图案的视觉感染力，更得益于文化叙事唤起的情感共鸣。如图 1 所示，国潮元素已深度融入家居、美妆等多个领域，形成视觉表达、故事内涵与品牌价值的有机统一[3]。然而，当前主流推荐算法多依赖单一模态或交互数据，难以解析这些分散且细腻的文化语义，也无法有效捕捉因文化认同而产生的动态偏好。因此，借助多模态深度学习技术，从多元数据中精准提炼文化特征以实现个性化推荐，已成为提升国潮消费体验的关键路径[4]。



Figure 1. Examples of the application of domestic trend cultural symbols across different product carriers
图 1. 国潮文化符号在不同产品载体中的应用示例

与此同时，推荐系统正从静态预测向动态交互演进。融合用户实时反馈的对话式推荐，不仅能提升精准度，更是构建深度品牌关系的重要桥梁，相关方法已展现出显著潜力[5]。为应对上述挑战，近年来多模态推荐研究逐渐从简单特征拼接转向更结构化的表征学习与跨模态对齐建模，并围绕模态选择、特征处理与融合机制展开系统探索。本文旨在系统梳理多模态推荐技术面向国潮文化场景的研究进展。结合国潮推荐的关键处理环节，从多模态表征学习、跨模态融合与对齐、图结构建模与知识增强、模型优化与训练策略四个维度对现有方法进行归纳，并分析各环节在国潮场景下的适配性，在此基础上，分析大语言模型驱动的生成式推荐范式在深层文化知识推理与可信解释生成方面的作用，最后探讨当前面临的挑战与未来发展方向，以期后续研究提供参考与借鉴。

2. 数据集

为全面、系统地评估与研究多模态推荐算法，尤其是在国潮品牌这一特定场景下的应用效果，选择与构建合适的数据集至关重要。多模态推荐研究通常需要同时建模用户行为、商品视觉信息、文本语义与结构化属性，以刻画用户的审美取向与文化认同。因此，本文选取并总结了当前电商多模态推荐领域中具有代表性的数据集，如表 1 所示。

Table 1. Summary table of multimodal recommendation datasets
表 1. 多模态推荐数据集汇总表

| 数据集 | 规模 | 模态 | 交互与标注 | 年份 | 场景与语言 |
|---------------------|---------|--------------|-----------|------|------------------|
| MEP-3M | 300 万商品 | 图像 标题 描述 类目 | 商品多模态信息 | 2022 | 中文电商 高质量标注 |
| Amazon Reviews 2023 | 数千万评论 | 图像 评论 评分 元数据 | 评分与评论 | 2023 | 国际电商 评论主观性强 |
| M5Product | 500 万商品 | 图像 标题 属性 | 商品多模态信息 | 2021 | 中文电商 商品信息丰富 |
| EMMa | 280 万对象 | 图像 标题 描述 属性 | 材料与属性 多标签 | 2023 | Amazon 商品 材料语义理解 |

续表

| | | | | | | | | |
|--------------|------|----------|---------------------|--------------|----------------|----------------------|------|-----------------|
| SIGIR eCom | 2021 | 493 万会话 | 3608 万事件 | 商品图像 会话行为 | 文本 价格 查询 | 会话点击与 购买 | 2021 | 搜索推荐 强会话场景 |
| H&M Fashion | | 3179 万交易 | 137 万用户 10.6 万商品 | 交易日志 文本属性 | 商品图像 用户画像 | 购买序列与 用户画像 | 2022 | 时尚电商 行为与内容结合 |
| UserBehavior | | 1 亿行为 | | 行为序列 | | 点击 收藏 加购 购买 | 2017 | 淘宝匿名行为 序列推荐 |

2.1. MEP-3M

MEP-3M(图 2)是一个大规模高质量的中文电商多模态数据集。其核心优势在于其高度结构化与精细化的数据组织。它不仅包含超过 300 万件商品的图像、标题文本、描述信息及层级化类目标签[6]，还并已按“配饰”、“珠宝”、“户外用品”等品类系统归类，为细粒度视觉表征学习、跨模态语义对齐和品类敏感的特征预训练而设计的系统性基准[7]。在国潮多模态推荐研究中，利用此类结构化数据预训练模型，可更精准地识别“户外”大类下的“国风露营”风格或“珠宝”品类中的“东方美学”元素。



Figure 2. MEP-3M dataset
图 2. MEP-3M 数据集

2.2. Amazon Reviews 2023



Figure 3. A detailed example entry from Amazon Reviews 2023 dataset
图 3. Amazon Reviews 2023 数据集中的一个详实例条

Amazon Reviews 数据集是国际电商推荐研究的常用基准，包含商品信息、用户评分及评论文本。

McAuley 等(2015)首次系统探索了图像与评论的联合建模，为多模态电商推荐奠定基础。其评论文本(图 3)涵盖材质、设计、场景及情感等深度描述，是语义挖掘与细粒度偏好提取的重要资源。该方法虽面向西方文化，但可迁移至国潮场景，用于从评论中提取“文化符号认同”、“质量满意度”等维度信号。

2.3. M5Product

M5Product 数据集(图 4)由 Cui 等于 2021 年发布，是一个面向商品多模态预训练任务的大规模数据集，包含约 500 万件淘宝商品的图像、标题文本及结构化属性信息[8]。其聚焦于商品侧多模态表征，不含用户交互数据，旨在为商品特征学习与跨模态任务提供预训练素材。商品图像类目广、质量高，标题简洁且含关键属性描述。在国潮多模态推荐研究中，可利用 M5Product 预训练视觉-语言模型，以编码商品的视觉风格、文化符号与语义信息。

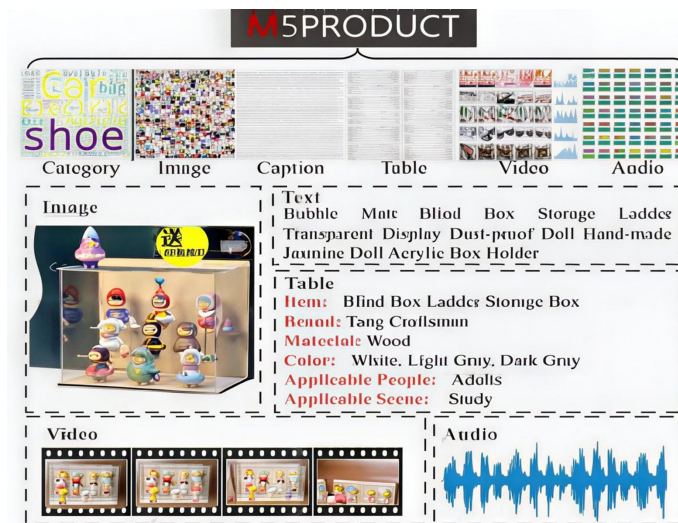


Figure 4. M5Product dataset

图 4. M5Product 数据集

2.4. EMMA

An Extensible Multimodal Multi-task Object Dataset with Materials

Trevor Standley¹ Ruohan Gao¹ Dawn Chen² Jiajun Wu¹ Silvio Savarese^{1,3}

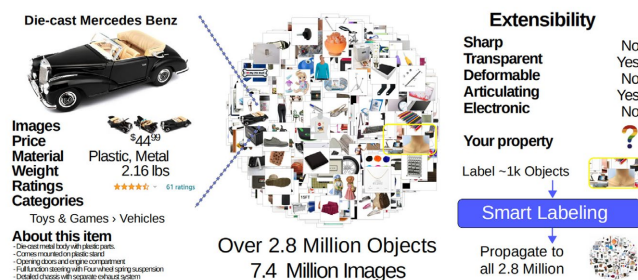


Figure 5. EMMA dataset

图 5. EMMA 数据集

EMMA 数据集(图 5)由 Standley 等于 2023 年提出，专注于电商商品的材料理解任务，涵盖超过 280

万件商品，每项包含商品图像、详情文本及结构化属性[9]。其核心贡献在于构建了包含 182 种材料的层级分类体系，并为商品标注多材料标签，有力支撑材料多标签分类、属性增强与多任务学习。在国潮检索与匹配任务中，EMMa 可作为属性增强数据源，借助材料监督引导模型学习“材质 - 外观 - 语义”的内在关联，同时为推荐结果提供清晰、可解释的属性依据。

2.5. 其他数据集

除上述详细介绍的数据集外，还有多个在特定方面具有研究价值的数据集。SIGIR eCom 2021 Data Challenge 数据集[10]由 Coveo 发布，提供真实的电商搜索与推荐会话日志，包含商品目录、用户查询，以及会话内完整的行为序列，例如点击、加购与购买等，是研究实时交互与序列化推荐的强场景基准；H&M Personalized Fashion Recommendations 数据集[11]融合了详尽的交易历史、用户属性与商品多模态信息，特别适用于时尚领域的个性化推荐与用户长期兴趣建模；UserBehavior 数据集[12]由阿里巴巴于 2017 年公开，包含约 1 亿条淘宝用户隐式反馈行为记录，其规模大、行为类型完整，被广泛用于序列推荐与用户兴趣演化建模。上述数据集在任务场景、数据模态和行为粒度上各有侧重，共同为构建精准、可解释且动态响应的国潮推荐系统提供了多元化的研究视角与评估基础。

3. 多模态推荐方法分类及其在国潮文化场景中的适用性分析

围绕国潮电商场景，多模态信息通常来自商品图像、评论文本、用户行为序列、文化知识图谱等。为系统利用这些异构数据，现有研究已形成多种方法框架。例如，Liu 等在其综述中从技术角度将多模态推荐模型归纳为四大类：模态编码器、特征交互、特征增强和模型优化[13]。这一分类涵盖了从特征提取、融合、语义增强到训练优化的完整技术链条，为分析多模态推荐方法提供了通用基础。

然而，面向国潮文化偏好这一特殊场景，我们需要进一步探讨这些通用技术如何适配文化符号识别、审美语义对齐、隐性文化关联挖掘等需求。因此，本文在 Liu 等分类的基础上，结合国潮推荐的关键处理环节，将代表性研究归纳为四个更聚焦的维度：多模态表征学习、跨模态融合与对齐、图结构建模与知识增强、模型优化与训练策略。图 6 展示了本文构建的面向国潮文化偏好的多模态推荐方法分类框架，从上述四个维度对现有研究进行系统梳理与适配性分析。以下将从这四个维度出发，梳理常见方法技术，并分析各环节在国潮场景下的研究进展与适配性。

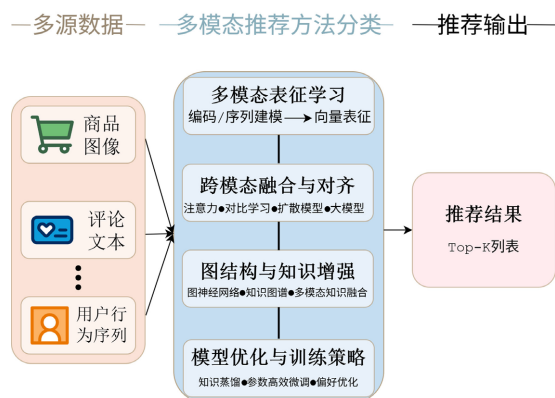


Figure 6. A classification framework of multimodal recommendation methods for domestic trend cultural preferences
图 6. 面向国潮文化偏好的多模态推荐方法分类框架

3.1. 多模态表征学习

多模态表征学习的核心目标是将图像、文本等异构数据转换为机器可理解的向量表示，为后续的融

合与推理提供基础。在国潮电商场景中，这一步骤需要让模型能够有效捕捉商品图像中的传统纹样、色彩体系等视觉符号，同时理解商品描述、用户评论中的文化叙事与审美意境。从技术演进脉络来看，多模态表征学习经历了从通用单模态编码器到面向推荐的任务自适应表征，再到大模型驱动的统一语义理解的发展过程。

早期研究主要借用计算机视觉和自然语言处理领域的预训练模型作为特征提取器，例如在视觉编码方面多采用 ResNet [14]、ViT [15]等模型提取特征，文本信息则常借助 BERT [16]等预训练语言模型进行编码。这些通用编码器虽为多模态推荐奠定基础，但其提取的特征并非为推荐任务专门设计，直接使用可能包含大量无关内容给后续推荐模型带来数据稀疏问题。为克服通用编码器的局限性，研究者提出了多种面向推荐任务优化的表征学习方法。如 Liu 等提出的 SGFD [17]采用教师-学生框架，教师模型从通用模态特征中提取丰富的语义信息和多模态互补信息，再通过响应蒸馏与特征蒸馏将知识迁移至学生模型，从而在缓解数据稀疏问题的同时，获得更贴合推荐任务的特征表示。进一步地，多模态大语言模型 (MLLM)的兴起为表征学习带来新的突破。Ye 等提出的 MLLM-MSR [18]模型探索了如何赋予 MLLM 多模态推荐能力：首先利用基于 MLLM 的物品总结器将图像转换为文本描述，再基于 LLM 的用户总结器以循环方式捕捉用户偏好的动态演化，最后通过监督微调使模型适应多模态推荐任务。该方法将视觉符号与文化语义统一在同一语义空间中，为跨模态对齐提供了新路径。

3.1. 国潮场景下的适配挑战与改进思路

在国潮场景下，现有多模态编码器在表征学习阶段面临着从“表层特征提取”向“深层文化符号识别”跨越的适配挑战。在视觉侧，国潮产品的核心特征常表现为特定的传统纹样与色彩体系，而通用视觉编码器缺乏对中国传统美学元素的先验知识，难以精准区分细粒度的文化语义。针对这一问题，改进思路在于利用特定领域的知识进行模型微调。例如，可构建包含大量中国传统纹样与细粒度文化描述的图文数据集，对 CLIP [19]模型进行视觉-语言对比微调。在微调过程中，可引入难负样本，如将视觉相似但文化内涵不同的其他地域纹样作为负样本，构建对比损失，并结合参数高效微调(Parameter-Efficient Fine-Tuning, PEFT) [20]技术，使模型在保留泛化能力的同时，具备精准识别和区分不同历史时期美学风格的能力。

在文本侧，国潮产品的描述与评论蕴含大量历史典故、文化叙事与审美意境，通用语言模型难以将其与用户的深层文化认同建立有效映射。对此，可通过文化知识注入的方式，利用外部知识图谱将文本中的文化实体映射为对应的文化向量。同时，可借助大语言模型的语义推理能力，将非结构化的文化叙事预先提炼为结构化的审美标签，从而有效弥补通用模型在国潮语境下的语义缺失，为后续的跨模态对齐提供坚实的表征基础。

3.2. 跨模态融合与对齐

在获得各模态的向量表征后，核心问题是如何有效整合这些异构信息，弥合其语义鸿沟。对齐旨在建立跨模态的语义对应关系，使它们的表征占据共享空间；而融合则将这些对齐后的特征整合为统一的预测或嵌入[21]。这两个过程紧密耦合，共同决定多模态推荐的准确性。从技术演进看，该领域方法经历了从基于注意力的早期融合、基于图的模态内传播，到当前基于对比学习、扩散模型和大语言模型的精细化对齐与融合阶段。

早期的代表性工作通过设计精细的神经网络结构，捕捉用户在多层级上的隐式偏好。Chen 等提出的 ACF [22]模型是首个将注意力机制引入协同过滤的推荐模型，旨在解决多媒体推荐中物品级与组件级的隐式反馈问题。随后，图神经网络因其在处理关系结构上的优势而被广泛采用。Wei 等提出的 MMGCN [23]框架不再简单融合多模态特征，而是为视觉、听觉和文本等每个模态分别构建用户-物品二分图，通

过在各个模态图上进行信息传播与聚合，生成模态特定的用户和物品表示，以捕捉用户在不同模态上的细粒度偏好。

近期研究则致力于解决特征对齐过程中的信息丢失与噪声引入问题。Yuan 等提出的 CLAM [24]模型指出，直接在不同模态特征之间对齐可能导致模态特有信息丢失。为此，CLAM 引入间接对齐机制，以物品 ID 嵌入为语义锚点，将其与不同模态特征进行对比学习，使各模态特征在拉近表示距离的同时保留独有信息，并通过多任务学习缓解对稀疏交互数据的过度依赖。Xiu 与 Tong 提出的 DCAR-DM [25]模型则采用更精细的双层对齐框架：首先利用扩散模型增强交互数据，然后通过特征对齐层将各模态特征建模为高斯分布，以最小化分布间的均值与标准差来实现分布层面对齐；接着通过行为对齐层，利用真实的用户-物品交互行为进行对比学习，以纠正特征对齐过程中产生的语义偏差。

此外，大语言模型也被用于增强融合与对齐的效果。Ma 等提出的 ExplainRec [26]框架利用大语言模型进行多模态增强，融合视觉与文本内容以处理冷启动场景，并通过偏好归因调优生成可解释的推荐结果。这一方向展示了将大模型的语义理解能力与多模态对齐任务相结合的潜力。

总体而言，对齐策略已从早期的隐式学习发展到显式的、多层次的、分布式的精细化对齐；融合方式则从简单的特征拼接或加权，演进到基于图的消息传递、对比约束以及大模型驱动的语义统一。

在国潮场景下的适配挑战与改进思路

在国潮电商场景中，跨模态对齐面临的核心挑战在于“视觉美学”与“文本叙事”之间存在巨大的语义鸿沟。国潮商品的吸引力往往源于其传达的特定“意境”，这种高阶语义在传统对比学习框架下极易被淹没在基础属性对齐中。改进思路在于引入文化锚点辅助对齐，即在对比学习的损失函数中加入“文化风格一致性”约束项，强制模型在共享隐空间中具有相同文化底蕴的图文向量拉近。此外，可采用多粒度对齐策略，不仅在全局层面进行图文匹配，还需通过注意力机制实现“局部纹样识别”与“文本文化关键词”的细粒度交互，从而确保模型能够捕捉到文化叙事与视觉符号之间的深层映射关系，提升推荐系统对用户审美偏好的感悟能力。

3.3. 图结构建模与知识增强

如何挖掘用户与商品之间、商品与商品之间的高阶关联，是提升推荐质量的关键。图结构建模与知识增强技术通过将用户行为、商品属性乃至文化概念组织为图结构，显式建模实体间的复杂关系，为深入理解用户偏好提供了新视角。

早期研究主要探索如何将知识图谱作为辅助信息引入推荐系统。Wang 等提出的 RippleNet [27]模拟用户偏好在知识图谱上的传播过程：将历史点击物品作为种子，沿知识链接迭代扩展兴趣形成多跳“涟漪”，叠加响应以刻画用户对候选物品的偏好，从而自动发现潜在关联路径并增强可解释性。该方法为知识图谱增强推荐提供了有效范式，但其主要利用结构化知识进行偏好传播，忽略了图像、文本等多模态信息。为此，Sun 等提出的 MKGAT [28]首次将多模态知识图谱引入推荐，设计多模态图注意力技术，通过不同编码器处理图像、文本等实体，在图注意力层聚合邻居时考虑关系类型，在保留推理关系的同时将多模态信息聚合到实体表示中。

在多模态融合的基础上，后续研究进一步聚焦于用户兴趣的个性化建模与图结构的精细化优化。针对用户对不同模态关注度不一致的问题，Wang 等提出的 DualGNN [29]构建双图神经网络：通过各模态的二分图捕捉单模态偏好，并利用用户共现图协同学习个性化融合权重，从而更精准地刻画用户的多模态兴趣。针对隐式反馈中普遍存在的误点击噪声问题，Wei 等提出的 GRCN [30]设计图细化层，通过原型网络学习用户内容偏好，自适应识别并修剪噪声边，再在优化图上进行图卷积，显著提升表示鲁棒性。

针对图卷积网络固有的过平滑问题, Ping 等提出的 Grade [31]从生成式对比学习角度进行缓解, 结合变分图自编码器与对比学习, 通过生成式图对比任务对齐模态特征, 并利用特征扰动对比任务增强模态表示的鲁棒性。Zhang 等提出的 MHMA-KGRec [32]则进一步深化多模态知识图谱的利用, 采用多头混合注意力机制实现跨关系与跨模态的细粒度特征交互, 结合跨模态对比学习与多模态决策融合, 增强模态间的语义对齐与互补利用。

图结构建模技术从基础图卷积发展到自适应图优化与生成式对比学习; 知识增强从单一知识图谱演进到多模态深度融合。

在国潮场景下的适配挑战与改进思路

现有的多模态知识图谱多侧重于建模商品与品牌、类目等客观属性的关联, 缺乏对国潮特有的“文化符号-情感意向”逻辑的建模, 难以挖掘用户对特定传统元素的隐性文化偏好。改进思路在于构建“中国文化符号-情感”异构图谱, 将“如意纹”、“莫兰迪色系”等抽象符号作为核心节点, 并建立起这些节点与“吉祥”、“素雅”等情感语义的关联边。在算法层面, 可利用图神经网络 GNN [33]的传播机制, 将用户的历史交互行为映射到该异构图谱中, 通过捕捉用户对特定文化元素的交互强度, 实现从“商品推荐”向“文化认同匹配”的升华。这种基于知识增强的建模方式, 不仅能显著提升推荐结果在长尾国潮商品上的准确性, 还能为推荐结果提供“因符合某种传统审美逻辑”的文化可解释性。

3.4. 模型优化与训练策略

多模态推荐模型的性能不仅取决于架构设计, 还深受优化策略与训练方式的影响。如何在多模态场景下平衡不同模态的贡献、缓解数据稀疏问题、提升训练效率与模型鲁棒性, 是模型优化环节需要解决的关键问题。

在多模态融合中, 不同模态的信息量往往不均衡, 统一优化目标易导致弱模态欠优化。Zhang 等提出的模态平衡学习方法针对这一问题, 采用反事实知识蒸馏技术, 通过单模态教师引导学生模型学习模态特定知识, 并利用反事实推断估计各模态对训练目标的因果效应, 自适应地重加权蒸馏损失, 使模型聚焦于弱模态[34]。针对多模态基础模型适配过程中参数效率与训练速度的权衡问题, Fu 等提出的 IISAN-Versa [35]框架采用解耦的参数高效微调结构, 通过组层丢弃与维度变换对齐策略有效处理非对称编码器, 在兼顾 GPU 内存效率的同时实现跨模态适配。

随着大语言模型引入推荐系统, 优化策略进一步向偏好对齐与多任务协同演进。Wang 等提出的 HaNoRec [36]框架针对监督微调后直接偏好优化中存在的样本难度不均衡与跨模态语义偏差问题, 设计难度感知重加权策略, 根据样本难度与模型实时响应动态调整优化权重, 并通过高斯扰动分布优化增强模态语义一致性。上述研究表明, 模型优化正从单一的损失函数设计向难度感知、模态平衡等多维度协同优化方向发展, 为国潮推荐中文化偏好建模与冷启动问题提供了新的解决思路。

在国潮场景下的适配挑战与改进思路

针对前文所述的多维度协同优化趋势, 国潮推荐场景下的核心适配挑战在于跨模态信息的严重失衡以及文化偏好建模中的数据稀疏问题。由于国潮商品高度依赖视觉美学传达文化意蕴, 而用户历史行为数据往往对价格、品牌等显式文本属性更为敏感, 导致模型在统一优化时易忽略深层文化特征的贡献。改进思路在于引入前文论述的反事实推断或自适应重加权机制, 在训练过程中动态补偿视觉模态中弱势文化符号的权重, 防止模型退化为纯文本驱动的逻辑。此外, 针对国潮领域品牌更迭快、新品比例高导致的冷启动挑战, 可进一步应用前文论述的难度感知重加权优化方案: 通过识别并加大对具有独特文化标签但交互较少的“硬样本”的学习力度, 缓解因样本分布不均带来的跨模态语义偏差。通过将用户表

征解耦为“通用流行偏好”与“特定文化偏好”进行多任务协同训练，不仅能提升模型在数据受限情况下的鲁棒性，更能在优化目标中实现从“点击预测”向“文化价值匹配”的深度对齐，从而有效解决长尾国潮商品的精准分发问题。

4. 推荐系统评估指标与对比方式

4.1. 离线评估设置与对比原则

推荐系统的效果通常通过离线评估进行衡量，即在历史交互数据上训练模型，再在未见过的测试数据集上检验其推荐质量。为保证对比公平，常见做法是将数据划分为训练集、验证集与测试集，例如按照80%、10%、10%的比例划分，并在相同的数据划分、相同的候选集规模与相同的Top-K推荐长度下比较不同方法的结果[37]。需要强调的是，电商推荐场景往往存在显著的类别不平衡与长尾分布，仅使用单一指标容易产生偏差，因此通常采用“相关性指标 + 排序质量指标 + 覆盖类指标”的组合进行更全面的评估。

4.2. 评估指标选择与含义说明

为了兼顾“推荐是否命中”、“命中位置是否靠前”以及“是否能够覆盖更多长尾商品”等关键目标，本文采用以下几类指标进行评价[38]：

1) Top-K 准确性指标

Top-K 的准确性通常用精确率与召回率刻画。精确率强调推荐列表的“命中纯度”，即模型推出来的K个物品中有多少是真正与用户偏好一致的；召回率强调“找回能力”，即用户在测试集中真实感兴趣的物品中，有多少被模型成功纳入推荐列表。二者分别对应减少误推荐与减少漏推荐，在隐反馈电商数据中是最常用的基础指标之一。

2) 命中率指标

该指标关注的是“目标物品是否进入列表”，因此对用户体验较直观：只要命中就计入，不区分命中发生在第几位。由于命中率对位置不敏感，通常需要与排序类指标配合使用，才能同时反映“命中”与“排在前面”的差异。

3) 排序质量指标

当推荐列表命中多个物品时，仅知道“命中与否”还不够，还需要衡量命中物品在列表中的位置是否靠前。该指标会对靠前位置赋予更高权重，更贴近真实使用场景，因此在序列推荐与深度推荐模型对比中经常作为核心指标之一。

4) 覆盖类指标

该指标反映模型在全体用户推荐中“覆盖了多少不同的商品”，覆盖范围越大，说明模型越有能力将长尾商品暴露给用户，而不仅仅反复推荐热门商品。在智能商品推荐系统设计中，也将覆盖率作为与精确率、召回率并列的重要指标，用于衡量对长尾物品的覆盖能力。在强化学习推荐实验中，覆盖率通常与精确率、排序质量指标一起报告，用于观察策略模型是否在提升相关性的同时牺牲了商品多样性与曝光广度。

4.3. 对比方式与基线方法组织

为使对比更具解释性，基线方法可按建模思想分为三类[39]：

第一类是非个性化或传统协同过滤基线，例如热门推荐以及矩阵分解类方法，它们实现简单但容易偏向热门物品，通常作为最低对照。

第二类是序列建模推荐方法，用于刻画用户行为的动态变化规律，包括马尔可夫链与深度序列模型。相关综述与实现工作指出，马尔可夫链方法擅长短期依赖但对长期兴趣刻画不足，而注意力机制与深度网络更利于捕捉复杂交互模式。第三类是更强的表示学习或策略学习方法，例如引入预训练与结构增强的序列模型，以及强化学习推荐，它们往往同时关注相关性与全局策略收益，因此在多个指标上会呈现不同的权衡。

4.4. 代表性实验结果对比

为直观呈现不同推荐方法在离线评估中的差异，本文选取两组具有代表性的对比结果：其一是 Amazon Baby 数据集上的序列推荐方法对比，用于反映主流深度序列模型在命中能力与排序质量方面的提升；其二是 SelectedUB 数据集上的强化学习推荐对比，用于展示策略学习方法在相关性与覆盖范围之间的权衡特征。两个表分别对应不同建模范式下的典型实验结论，可为后续方法选择与实验设计提供参考依据[40][41]。

Table 2. Comparative results of sequential recommendation on the Amazon Baby dataset

表 2. Amazon Baby 数据集序列推荐对比结果

| 方法 | HR@1 | HR@5 | HR@10 | NDCG@5 | NDCG@10 |
|-----------|--------|--------|--------|--------|---------|
| FPMC | 0.0385 | 0.1571 | 0.2573 | 0.0954 | 0.1238 |
| GRU4Rec | 0.0534 | 0.2141 | 0.3381 | 0.1364 | 0.1841 |
| SASRec | 0.0885 | 0.2559 | 0.3783 | 0.1727 | 0.2147 |
| SSE-PT | 0.0937 | 0.2585 | 0.387 | 0.1776 | 0.2158 |
| CARCA | 0.0946 | 0.2595 | 0.3995 | 0.1794 | 0.2175 |
| CARCA-SSE | 0.0959 | 0.2697 | 0.4017 | 0.1842 | 0.2267 |

从表 2 可以看出，非个性化与传统协同过滤基线(FPMC)在各项指标上整体偏低；随着模型引入序列建模能力(GRU4Rec、SASRec)，HR 与 NDCG 均出现稳定提升，说明用户行为序列信息对于提升推荐质量具有关键作用。在此基础上，引入更强表示增强策略的模型(SSE-PT、CARCA、CARCA-SSE)进一步提高了命中率与排序质量，其中 CARCA-SSE 在 HR@10 与 NDCG@10 上取得最优结果，体现了更强结构建模对列表推荐效果的促进作用。

Table 3. Comparative results of reinforcement learning-based recommendation on SelectedUB

表 3. 强化学习推荐在 SelectedUB 上的对比结果

| 算法 | Top-10 Precision | Top-10 NDCG | Top-10 Coverage | Top-20 Precision | Top-20 NDCG | Top-20 Coverage |
|---------|------------------|-------------|-----------------|------------------|-------------|-----------------|
| ItemPop | 0.1 | 3.6 | 0.3 | 0.05 | 1.1 | 0.5 |
| DDPG | 0.2 | 0.8 | 0.002 | 0.1 | 0.9 | 0.003 |
| BCQ | 2.1 | 4.1 | 8.9 | 1.5 | 6.1 | 24.5 |
| NaPRS | 2.3 | 4.6 | 16.0 | 1.8 | 6.8 | 26.8 |

表 3 反映了策略类方法在不同目标上的差异：DDPG 与 NaPRS 在 Precision 与 NDCG 上更具优势，说明策略学习能够有效提升推荐相关性与排序质量；同时，BCQ 与 NaPRS 在 Coverage 指标上更高，表明其推荐结果覆盖的物品范围更广，对长尾曝光更友好。综合比较，NaPRS 在 Top-10 与 Top-20 的相关性指标与覆盖指标上均表现较优，体现出其在推荐准确性与覆盖范围之间相对更好的平衡能力。

5. 现有挑战

5.1. 文化语义难以被可计算化建模

国潮推荐需要刻画商品的文化符号与审美意涵，但其多以隐性形式分布在多模态中：视觉侧体现纹样、配色、材质质感等细节，文本侧多为意象叙事与情绪表达；同时统一稳定的文化标签与高质量监督稀缺，使文化语义难以结构化为可学习表征。现有方法因此易仅拟合表层相似，难以区分文化一致与外观相近，导致推荐“相关却不契合”[42]。此外，文化偏好与价格、热度、曝光等商业因素强耦合，若缺乏对混杂因素的显式控制，模型更可能学习平台机制而非真实文化认同，削弱泛化与可信度[43]。

5.2. 跨模态对齐容易产生对齐错位

国潮场景存在突出的语义层级不一致：图像侧偏细粒度符号与风格线索，文本侧承载联想、评价与叙事语义；若将两者简单映射到共享空间并以相似度对齐，容易把相关性误当等价性。叠加电商数据中的UGC噪声、营销化表述、标签粗糙与属性缺失，对齐偏差会在融合与排序中被放大，表现为视觉相似但文化语境不匹配的错配推荐[44]。多源视图的互补与冲突并存也会加剧融合不稳定，导致跨品类、跨平台迁移时性能波动。

5.3. 偏好动态变化导致推荐稳定性不足

国潮消费强时效、强传播，热点与季节场景持续驱动偏好漂移，使用户同时呈现长期审美取向与短期潮流冲动。现有动态建模难以兼顾快速响应与长期一致：过敏感易追热点并遗忘长期偏好，过平滑则错失潮流窗口。进一步地，多模态更新不同步带来时间错配与漂移累积，如内容变化快、属性更新滞后、反馈稀疏且延迟等，再叠加在线推理时延与增量更新成本等约束，离线有效策略难以稳定落地，推荐质量随时间波动[45]。

5.4. 解释难以做到可信且可验证

国潮推荐解释不仅要说明原因，更应提供可核验的文化证据，使决策依据可追溯到具体线索。现有解释多停留在概括性描述，难以保证与模型决策一致，易出现“解释强调文化、模型依赖热度”等脱节，损害信任[46]。同时文化概念边界模糊、语境依赖强，缺乏统一的解释标注体系与评测协议，解释质量难以标准化比较，也难以严谨评估其对信任、复购与口碑等长期目标的真实增益，限制工程复用价值。

6. 大语言模型驱动为国潮推荐：技术破局与范式革新

随着多模态大语言模型(MLLMs)的发展，推荐系统正从传统的特征拼接转向基于深度语义理解的生成式推荐。鉴于国潮商品的文化语义难以被可计算化建模，大语言模型凭借海量的参数化知识与强大的跨模态能力，为解决文化语义挖掘与可信解释生成提供了新路径[47][48]。

6.1. 基于大语言模型的深层文化知识推理

相较于依赖表层向量距离的传统对齐方法，大语言模型能够实现更深层次的文化知识推理：一是解码隐性语义，直接识别图像中的复杂文化符号，如传统纹样，并结合知识库推理其历史渊源[47]；二是跨模态消歧，通过逻辑比对过滤图文不一致的营销噪声，精准提取核心文化特征[48]；三是偏好动态推演，通过分析长序列交互与文本评论，推演用户审美演进路径，实现系统性的文化偏好预测[49]。这种深度推理能力为解决第5.1节中提到的文化语义难以被可计算化建模和第5.2节的跨模态对齐错位问题提供了有效途径。

6.2. 生成式可信国潮推荐解释

针对现有推荐解释难以做到可信且可验证的痛点，大语言模型能突破传统模板限制，融合多模态特征生成个性化、连贯的文化叙事解释。进一步地，通过引入检索增强生成(RAG)技术，系统可实时检索历史文献或博物馆公开数据作为上下文[50]。这种机制不仅能有效缓解大模型的“事实幻觉”，还能为推荐提供可追溯的文化证据，从而显著提升用户的文化认同感与对系统的信任度。此技术路线为第 5.4 节所述的解释可信性挑战提供了可行解决方案，使解释不再停留于表层描述，而是能够提供可验证的文化依据链。

6.3. 交互式对齐与偏好可控建模

面对主观性极强的国潮审美，大语言模型支持通过自然语言对话构建交互式推荐，获取用户的显式纠偏反馈。结合监督微调(SFT)与直接偏好优化(DPO)等技术，模型能更精准地适配人类的文化审美标准[51]。该机制不仅能敏捷响应短期的潮流波动，还能在多轮交互中锁定用户的长期文化价值观，实现推荐结果的动态稳定与精准可控。这种交互式对齐框架有效应对了第 5.3 节讨论的偏好动态变化与稳定性平衡难题，为国潮推荐系统提供了人机协同演化的技术路径。

7. 总结与未来趋势

7.1. 总结

国潮消费的关键不再只是“好用、便宜”，而是商品能否传递某种文化意味与审美态度。也正因为如此，国潮商品往往同时包含图像风格、文本叙事与属性标签等多种线索，用户的偏好也更细、更主观、更容易随情境变化。本文围绕国潮文化偏好的多模态电商推荐，梳理了数据形态与公开数据资源，并从表征学习、融合方式、跨模态关联、知识与图结构利用，以及序列行为与交互反馈等角度归纳代表性研究；此外，本文分析了大语言模型驱动的推荐范式，探讨其在文化语义挖掘与可信解释生成等方面的技术破局作用。总体来看，多模态方法确实为国潮推荐提供了更丰富的“证据来源”，但在文化语义表达、模态信息匹配、偏好变化建模与解释可信度等方面，仍存在明显的研究空白与落地难点。特别是，大语言模型为解决这些难点提供了新思路，通过深层文化知识推理、生成式可信解释和交互式偏好对齐，有望突破传统推荐范式的局限。

7.2. 未来趋势

总体而言，国潮多模态推荐的后续发展重点不在于继续增加输入信息的种类，而在于把已有信息用得更准确、更稳定、更可信。一方面，国潮语境中的文化内涵往往以隐含方式存在于图像风格细节、文案表达与结构化属性之中，未来研究需要更好地将这些文化线索转化为可计算、可复用的语义表示，使模型能够捕捉“文化相关”而非仅停留在“表面相似”。大语言模型与多模态基础模型的结合将成为这一转化的关键技术路径。另一方面，电商场景下图像与文本之间并不总是严格对应，营销表达、描述侧重点差异等因素会带来信息不一致甚至冲突，未来更应强调跨模态关联的可靠性，避免将不匹配的信息强行融合从而放大误差。大语言模型的语义消歧与逻辑推理能力，为提升跨模态关联可靠性提供了新工具。此外，国潮消费具有明显的时效性与场景性，用户既存在相对稳定的审美倾向，也会受到热点、联名与节日情境的短期影响，因此如何同时刻画长期偏好与短期变化，并在更新过程中保持推荐结果的稳定性，将成为影响落地效果的重要问题。交互式大语言模型与用户偏好动态建模的结合，有望在响应短期波动与保持长期一致性之间取得平衡。最后，国潮推荐的解释应更贴近用户的理解方式，并尽量与模型决策保持一致，通过清晰的依据链路增强可验证性与信任度。基于检索增强的大语言模型，能够为推

荐提供可追溯的文化证据,使解释从“模型认为”转变为“有据可查”,从而显著提升用户对系统的信任度。随着多模态模型在电商中的应用持续深化,围绕文化语义表达、跨模态一致性、偏好动态刻画与可信解释的系统化改进,尤其是大语言模型驱动的技术创新,将是该方向最具价值的研究趋势。

基金项目

2025年贵州大学创新创业训练计划项目(项目编号:gzugc2025012)。

参考文献

- [1] 张时俊,王永恒.基于矩阵分解的个性化推荐系统研究[J].中文信息学报,2017,31(3):134-139,169.
- [2] 袁文华.国家认同视域下青年国潮消费的特征、动因与引领[J].中国青年研究,2024(11):4-11,94.
- [3] 唐卓亚,杨娟.青年国潮消费背后的文化自信与价值引导研究——以电商平台为例[J].电子商务评论,2025,14(12):3848-3856.
- [4] 郝雅娴,孙艳蕊.K-近邻矩阵分解推荐系统算法[J].小型微型计算机系统,2018,39(4):755-758.
- [5] Li, H., Huang, X., Tian, W. and Chen, X. (2026) Causal Interest Modeling and Popularity Bias Mitigation in Conversational Recommender Systems. *Knowledge-Based Systems*, **331**, Article ID: 114806. <https://doi.org/10.1016/j.knosys.2025.114806>
- [6] Liu, F., Chen, D., Du, X., Gao, R. and Xu, F. (2023) MEP-3M: A Large-Scale Multi-Modal E-Commerce Product Dataset. *Pattern Recognition*, **140**, Article ID: 109519. <https://doi.org/10.1016/j.patcog.2023.109519>
- [7] Chen, S., Bouadjenek, M.R., Jameel, S., Naseem, U., Suleiman, B., Salim, F.D., Hacid, H. and Razzak, I. (2025) Leveraging Taxonomy and LLMs for Improved Multimodal Hierarchical Classification. *Proceedings of the 31st International Conference on Computational Linguistics (COLING 2025)*, Abu Dhabi, 19-24 January 2025, 6244-6254.
- [8] Cui, Y., Liu, Y., Liu, X., Wang, Y. and Zhu, Y. (2021) M5Product: A Large-Scale Multimodal Product Dataset. arXiv: 2109.04275.
- [9] Standley, T.S., Gao, R., Chen, D., Wu, J. and Savarese, S. (2023) An Extensible Multi-Modal Multi-Task Object Dataset with Materials. *The Eleventh International Conference on Learning Representations (ICLR 2023)*, Kigali, 1-5 May 2023, 1-18. <https://openreview.net/forum?id=n70oy1IS4g>
- [10] Gupta, A., Mehrotra, R., Bhattacharya, P., Sharma, A. and Chandar, P. (2021) The SIGIR 2021 eCom Data Challenge. *SIGIR Forum*, **55**, Article 19.
- [11] H&M Group (2022) H&M Personalized Fashion Recommendations. Kaggle. <https://www.kaggle.com/competitions/h-and-m-personalized-fashion-recommendations>
- [12] Zhu, H., Chang, D., Xu, Z., Zhang, P., Li, X., He, J., et al. (2024) UserBehavior: A Dataset for Recommendation. Alibaba Cloud Tianchi. <https://tianchi.aliyun.com/dataset/dataDetail?dataId=649>
- [13] Liu, Q., Hu, J., Xiao, Y., Zhao, X., Gao, J., Wang, W., et al. (2024) Multimodal Recommender Systems: A Survey. *ACM Computing Surveys*, **57**, 1-17. <https://doi.org/10.1145/3695461>
- [14] He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep Residual Learning for Image Recognition. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 27-30 June 2016, 770-778. <https://doi.org/10.1109/cvpr.2016.90>
- [15] Dosovitskiy, A., Beyer, L., Kolesnikov, A., et al. (2020) An Image Is Worth 16 × 16 Words: Transformers for Image Recognition at Scale. arXiv: 2010.11929.
- [16] Devlin, J., Chang, M.W., Lee, K., et al. (2019) Bert: Pre-Training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, Minneapolis, 2-7 June 2019, 4171-4186.
- [17] Liu, F., Chen, H., Cheng, Z., Nie, L. and Kankanhalli, M. (2023) Semantic-Guided Feature Distillation for Multimodal Recommendation. *Proceedings of the 31st ACM International Conference on Multimedia*, Ottawa, 29 October-3 November 2023, 6567-6575. <https://doi.org/10.1145/3581783.3611886>
- [18] Ye, Y., Zheng, Z., Shen, Y., Wang, T., Zhang, H., Zhu, P., et al. (2025) Harnessing Multimodal Large Language Models for Multimodal Sequential Recommendation. *Proceedings of the AAAI Conference on Artificial Intelligence*, **39**, 13069-13077. <https://doi.org/10.1609/aaai.v39i12.33426>
- [19] Radford, A., Kim, J.W., Hallacy, C., et al. (2021) Learning Transferable Visual Models from Natural Language Supervision.

- International Conference on Machine Learning*. PmLR, 2021, 18-24 July 2021, 8748-8763.
- [20] Ding, N., Qin, Y., Yang, G., Wei, F., Yang, Z., Su, Y., *et al.* (2023) Parameter-Efficient Fine-Tuning of Large-Scale Pre-Trained Language Models. *Nature Machine Intelligence*, **5**, 220-235. <https://doi.org/10.1038/s42256-023-00626-4>
- [21] Li, S. and Tang, H. (2024) Multimodal Alignment and Fusion: A Survey. arXiv: 2411.17040.
- [22] Chen, J., Zhang, H., He, X., Nie, L., Liu, W. and Chua, T. (2017) Attentive Collaborative Filtering: Multimedia Recommendation with Item- and Component-Level Attention. *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Shinjuku, 7-11 August 2017, 335-344. <https://doi.org/10.1145/3077136.3080797>
- [23] Wei, Y., Wang, X., Nie, L., He, X., Hong, R. and Chua, T. (2019) MMGCN: Multi-Modal Graph Convolution Network for Personalized Recommendation of Micro-Video. *Proceedings of the 27th ACM International Conference on Multimedia*, Nice, 21-25 October 2019, 1437-1445. <https://doi.org/10.1145/3343031.3351034>
- [24] Yuan, X., Qi, A., Wu, H., Wang, J., Guo, Y., Li, S., *et al.* (2025) Cross-Modal Feature Alignment and Fusion with Contrastive Learning in Multimodal Recommendation. *Knowledge-Based Systems*, **326**, Article ID: 114020. <https://doi.org/10.1016/j.knosys.2025.114020>
- [25] Xiu, Y. and Tong, X. (2026) Dual-Layer Cross-Modal Alignment Recommendation Based on the Diffusion Model. *Information Fusion*, **125**, Article ID: 103472. <https://doi.org/10.1016/j.inffus.2025.103472>
- [26] Ma, B., Liu, L.Y., Hu, Z.H., *et al.* (2025) ExplainRec: Towards Explainable Multi-Modal Zero-Shot Recommendation with Preference Attribution and Large Language Models. arXiv: 2511.14770.
- [27] Wang, H., Zhang, F., Wang, J., Zhao, M., Li, W., Xie, X., *et al.* (2018) RippleNet: Propagating User Preferences on the Knowledge Graph for Recommender Systems. *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, Torino, 22-26 October 2018, 417-426. <https://doi.org/10.1145/3269206.3271739>
- [28] Sun, R., Cao, X., Zhao, Y., Wan, J., Zhou, K., Zhang, F., *et al.* (2020) Multi-Modal Knowledge Graphs for Recommender Systems. *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 19-23 October 2020, 1405-1414. <https://doi.org/10.1145/3340531.3411947>
- [29] Wang, Q., Wei, Y., Yin, J., Wu, J., Song, X. and Nie, L. (2023) Dualgnn: Dual Graph Neural Network for Multimedia Recommendation. *IEEE Transactions on Multimedia*, **25**, 1074-1084. <https://doi.org/10.1109/tmm.2021.3138298>
- [30] Wei, Y., Wang, X., Nie, L., He, X. and Chua, T. (2020) Graph-Refined Convolutional Network for Multimedia Recommendation with Implicit Feedback. *Proceedings of the 28th ACM International Conference on Multimedia*, Seattle, 12-16 October 2020, 3541-3549. <https://doi.org/10.1145/3394171.3413556>
- [31] Ping, Y., Wang, S., Yang, Z., Dong, Y., Hu, M. and Zhang, P. (2025) Grade: Generative Graph Contrastive Learning for Multimodal Recommendation. *Neurocomputing*, **657**, Article ID: 131630. <https://doi.org/10.1016/j.neucom.2025.131630>
- [32] Zhang, S., Yang, L. and Cheng, Q. (2026) A Multi-Head Mixed Attention Mechanism Enhanced Multimodal Knowledge Graph for Personalized Recommendation. *Neurocomputing*, **667**, Article ID: 132393. <https://doi.org/10.1016/j.neucom.2025.132393>
- [33] Schlichtkrull, M., Kipf, T.N., Bloem, P., van den Berg, R., Titov, I. and Welling, M. (2018) Modeling Relational Data with Graph Convolutional Networks. In: Gangemi, A., *et al.*, Eds., *The Semantic Web*, Springer, 593-607. https://doi.org/10.1007/978-3-319-93417-4_38
- [34] Zhang, J., Liu, G., Liu, Q., Wu, S. and Wang, L. (2024) Modality-Balanced Learning for Multimedia Recommendation. *Proceedings of the 32nd ACM International Conference on Multimedia*, Melbourne, 28 October-1 November 2024, 7551-7560. <https://doi.org/10.1145/3664647.3680626>
- [35] Fu, J., Ge, X., Xin, X., Karatzoglou, A., Arapakis, I., Zheng, K., *et al.* (2025) Efficient and Effective Adaptation of Multimodal Foundation Models in Sequential Recommendation. *IEEE Transactions on Knowledge and Data Engineering*, **37**, 7076-7089. <https://doi.org/10.1109/tkde.2025.3608071>
- [36] Wang, Y., Yang, Y., Wu, L., *et al.* (2025) Multimodal Large Language Models with Adaptive Preference Optimization for Sequential Recommendation. arXiv: 2511.18740.
- [37] Zhou, H., Zhou, X., Zeng, Z., *et al.* (2023) A Comprehensive Survey on Multimodal Recommender Systems: Taxonomy, Evaluation, and Future Directions. arXiv: 2302.04473.
- [38] Zhang, S., Yao, L., Sun, A., *et al.* (2019) Deep Learning Based Recommender System: A Survey and New Perspectives. *ACM Computing Surveys (CSUR)*, **52**, 1-38.
- [39] 刘婷, 朱亚峰. 基于机器学习的智能商品推荐系统设计[J]. 中国新技术新产品, 2025(20): 8-11.
- [40] 徐昊栋. 基于深度强化学习的商品推荐系统[D]: [硕士学位论文]. 杭州: 杭州电子科技大学, 2025.
- [41] 曲照鑫. 基于深度学习的商品推荐算法研究与软件开发[D]: [硕士学位论文]. 沈阳: 沈阳工业大学, 2024.

-
- [42] Liu, Q., Hu, J., Xiao, Y., Zhao, X., Gao, J., Wang, W., *et al.* (2024) Multimodal Recommender Systems: A Survey. *ACM Computing Surveys*, **57**, 1-17. <https://doi.org/10.1145/3695461>
- [43] Joachims, T., Swaminathan, A. and Schnabel, T. (2017) Unbiased Learning-to-Rank with Biased Feedback. *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, Cambridge, 6-10 February 2017, 781-789. <https://doi.org/10.1145/3018661.3018699>
- [44] Liang, W., Zhang, Y., Kwon, Y., Yeung, S. and Zou, J. (2022) Mind the Gap: Understanding the Modality Gap in Multi-Modal Contrastive Representation Learning. arXiv: 2203.02053.
- [45] Kang, W. and McAuley, J. (2018) Self-Attentive Sequential Recommendation. 2018 *IEEE International Conference on Data Mining (ICDM)*, Singapore, 17-20 November 2018, 197-206. <https://doi.org/10.1109/icdm.2018.00035>
- [46] Chen, X., Zhang, Y. and Wen, J.R. (2022) Measuring “WHY” in Recommender Systems: A Comprehensive Survey on the Evaluation of Explainable Recommendation. arXiv: 2202.06466.
- [47] Liu, H., Li, C., Wu, Q. and Lee, Y.J. (2023) Visual Instruction Tuning. arXiv: 2304.08485.
- [48] Li, J., Li, D., Savarese, S. and Hoi, S. (2023) BLIP-2: Bootstrapping Language-Image Pre-Training with Frozen Image Encoders and Large Language Models. *Proceedings of the 40th International Conference on Machine Learning*, Honolulu, 23-29 July 2023, 19730-19742.
- [49] Deng, Y., Zhang, W., Xu, W., *et al.* (2023) LLM-Rec: Large Language Models for Sequential Recommendation. *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Toronto, 24-28 July 2023, 1512-1521.
- [50] Lewis, P., Perez, E., Piktus, A., *et al.* (2020) Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. arXiv: 2005.11401.
- [51] Rafailov, R., Sharma, A., Mitchell, E., *et al.* (2023) Direct Preference Optimization: Your Language Model is Secretly a Reward Model. arXiv: 2305.18290.