

SCMRec: 基于 SAC 与对比学习的多模态推荐

钱姜枚, 李唯唯

(重庆理工大学 计算机科学与工程学院, 重庆 400054)

摘要: 随着社交媒体的普及, 用户兴趣表达高度依赖多模态内容融合。然而, 现有推荐方法面临根本性局限, 如静态特征融合导致兴趣建模滞后, 监督学习依赖标注数据加剧长尾覆盖不足, 强化学习忽视多模态互补性引发决策偏差等, 严重降低了推荐的准确性与公平性。为此, 提出多模态强化学习推荐框架 SCMRec, 将 Soft Actor-Critic 与多模态动态交互建模结合, 并在策略网络中引入对比学习模块。具体来说, 基于强化学习 Soft Actor-Critic 算法构建实时更新的用户-物品交互图状态表示, 解决传统方法静态特征融合难以捕捉用户兴趣时序演化的问题, 实现了细粒度兴趣动态追踪; 在 Actor 网络中嵌入对比学习模块, 通过最大化跨模态正负样本表征距离来解决稀疏性问题; 同时设计联合优化即时奖励与多样性目标的损失函数来抑制热门偏差。在 Amazon-Baby、Sports 和 Allrecipes 数据集上将 SCMRec 与 8 种基线方法进行比较, 实验结果表明, SCMRec 的 Recall@20 和 NDCG@20 指标超越最优基线, 且在极端稀疏场景下表现尤为突出。消融实验验证了动态交互图、Soft Actor-Critic 算法与对比学习的协同效应。该研究为多模态推荐系统提供了动态建模新范式, 推动了强化学习与对比学习的协同应用。

关键词: 多模态; 推荐系统; SAC 算法; 强化学习; 对比学习

DOI: 10.11907/rjtk.251583

扫描二维码阅读全文: 

中图分类号: TP391

文献标识码: A

文章编号: 1672-7800(XXXX)0XX-0001-08

SCMRec: Multimodal Recommendation Based on SAC and Contrastive Learning

QIAN Jiangmei, LI Weiwei

(College of Computer Science and Engineering, Chongqing University of Technology, Chongqing 400054, China)

Abstract: With the widespread adoption of social media, users' expression of interests has become highly reliant on the fusion of multimodal content. However, existing recommendation methods face fundamental limitations. For instance, static feature fusion leads to lagging interest modeling, supervised learning's reliance on labeled data exacerbates inadequate coverage of long-tail items, and reinforcement learning's neglect of multimodal complementarity triggers decision biases, all of which significantly degrade the accuracy and fairness of recommendations. To address these issues, we propose a multimodal reinforcement learning recommendation framework named SCMRec, which integrates Soft Actor-Critic with multimodal dynamic interaction modeling and incorporates a contrastive learning module into the policy network. Specifically, SCMRec constructs a real-time updated state representation of the user-item interaction graph based on the Soft Actor-Critic algorithm in reinforcement learning. This approach resolves the challenge of traditional methods' static feature fusion, which struggles to capture the temporal evolution of user interests, enabling fine-grained dynamic tracking of interests. A contrastive learning module is embedded in the Actor network to address sparsity issues by maximizing the representation distances between cross-modal positive and negative samples. Additionally, a loss function that jointly optimizes immediate rewards and diversity objectives is designed to suppress popularity bias. Comparative experiments on the Amazon-Baby, Sports, and Allrecipes datasets, where SCMRec is evaluated against eight baseline methods, demonstrate that SCMRec outperforms the best baseline in terms of Recall@20 and NDCG@20 metrics, particularly excelling in extremely sparse scenarios. Ablation experiments validate the synergistic effects of the dynamic interaction graph, the Soft Actor-Critic algorithm, and contrastive learning. This

收稿日期: 2025-08-24

作者简介: 钱姜枚 (2001-), 女, 重庆理工大学计算机科学与工程学院硕士研究生, 研究方向为智能信息推荐; 李唯唯 (1976-), 女, 重庆理工大学计算机科学与工程学院副教授、硕士生导师, 研究方向为大数据技术及应用、数据分析、智能信息推荐。本文通讯作者: 钱姜枚。

study offers a novel paradigm for dynamic modeling in multimodal recommendation systems and advances the collaborative application of reinforcement learning and contrastive learning.

Key Words: multimodal; recommender systems; SAC algorithms; reinforcement learning; contrastive learning

0 引言

随着移动互联网与多媒体技术的快速发展,用户生成内容呈现爆炸式增长。个性化推荐系统作为建立用户偏好与海量资源之间精准映射的核心技术,已广泛应用于电子商务、社交分发及在线娱乐等多个领域^[1]。随着用户交互载体从单一的标识符扩展至丰富的图像、文本等异构数据,多模态推荐通过整合多维语义信息,显著增强了系统对复杂物品属性的刻画能力,成为提升分发效率的关键路径^[2]。然而,现有推荐范式在实际应用中仍面临严峻挑战。首先,传统方法多侧重于静态建模,难以实时捕捉用户偏好的动态演化特征^[3];其次,在交互数据极度稀疏的场景下,高维特征的有效转化与判别性增强仍是该领域的理论难点^[4];最后,如何在维持推荐精准性的前提下兼顾结果的多样性并抑制热门偏差,直接影响到平台生态的长期健康。因此,探索具备高度动态感知力、表征稳健性及多目标权衡能力的推荐算法对于优化用户长程交互体验、推动推荐技术向智能化决策范式演进具有重要的学术价值与实践意义。

1 相关工作

推荐系统的核心任务在于精准捕捉用户偏好并预测其潜在兴趣。随着互联网多媒体内容的爆炸式增长,传统协同过滤(Collaborative Filtering, CF)方法在应对高稀疏性、冷启动等挑战时显得力不从心^[5],这促使研究者致力于融合多模态信息(如视觉、文本等)以更全面地表征项目特征与用户意图^[2,6]。与此同时,将推荐过程序列化建模为动态决策问题的强化学习方法可显式捕捉用户兴趣演化,因而受到学术界广泛关注^[7-8]。以下围绕多模态推荐与强化学习推荐两大前沿方向对现有研究进行系统梳理。

1.1 多模态推荐

多模态推荐系统通过整合视觉、文本等多维度信息来提升推荐性能,已成为当前研究的热点。根据技术路线的不同,现有研究可分为以下3类:①将多模态特征融入推荐框架。例如,He等^[9]提出的VBPR模型将视觉特征与用户嵌入向量相结合,扩展了贝叶斯个性化排序(Bayesian Personalized Ranking, BPR)框架,但仅通过简单拼接未能充分挖掘模态间的语义关联。Wei等^[10]提出的MMGCN模型为每个模态构建独立的图结构,利用图卷积网络聚合多跳邻居信息,显著提升了推荐性能,但其静态图结构难以适应动态变化的用户兴趣。随后,其引入图细化机制动态优化图结构,但仍受限于固定的图传播范式^[11];②采用自

监督学习方法解决数据稀疏性问题。例如,Tao等^[12]提出的SLMRec通过执行特征掩蔽与模态模式发现,从物品的图文属性中提取细粒度语义约束可在交互数据极少的情況下学习到高质量的物品表征。Wei等^[13]提出的MMSSL通过构建模态感知的交互结构学习任务,利用跨模态的一致性对齐增强了模型在稀疏图拓扑下的鲁棒性,有效弥补了用户行为信号的缺失。虽然该类方法通过自监督机制强化了表征深度,但在优化过程中仍高度依赖BPR等传统损失函数,由于其天然偏好高频交互数据,模型依然容易向热门物品偏移,难以从根本上解决长尾项目的推荐公平性问题^[14];③采用跨模态对齐方法实现不同模态语义空间的对齐^[15-16]。例如,Jiang等^[17]提出的DiffMM模型采用多模态扩散技术生成用户—项目交互图,并结合对比学习实现特征对齐。然而,这类方法对高质量对齐数据依赖性强,且缺乏对时序动态性的充分考虑。

1.2 强化学习推荐

强化学习为推荐系统提供了建模用户兴趣动态演化的新途径,其发展历程呈现出清晰的时序演进脉络。在早期研究中,Watkins等^[18]提出的Q-learning算法奠定了值函数方法的理论基础。Kaelbling等^[7]发表的强化学习综述系统梳理了该领域的核心理论与方法,为后续研究提供了重要参考。Shani等^[8]开创性地将推荐问题形式化为马尔可夫决策过程(Markov Decision Process, MDP),为强化学习在推荐系统中的应用奠定了框架基础。随着深度学习的兴起,强化学习与深度神经网络逐渐相结合。例如,Mnih等^[19]提出的DQN(Deep Q-Network)算法将深度神经网络与Q-learning相结合,解决了高维状态空间的表示问题。Lillicrap等^[20]提出深度确定性策略梯度算法,为连续动作空间中的策略优化提供了有效解决方案。Zhao等^[21]将经验回放机制引入深度强化学习算法SARSA中,显著提升了该算法解决复杂控制问题的能力。Wei等^[22]将强化学习应用于排序问题,提出基于MDP的排序框架,为推荐系统的排序优化提供了新思路。Zheng等^[23]将DQN算法应用于新闻推荐,提出DRN框架,显式地建模用户兴趣的动态变化。Haarnoja等^[24]提出的SAC(Soft Actor-Critic)算法是一种基于最大熵强化学习框架的深度强化学习算法,在连续控制领域展现出优异性能。Chen等^[25]提出的TPGR算法通过树结构策略梯度有效处理大规模离散动作空间,为解决推荐系统中的动作空间爆炸问题提供了创新方案。近年来,强化学习推荐系统向更广泛的应用领域和更精细的算法设计方向发展。例如,Liu等^[26]将强化学习应用于食品推荐系统;Luo等^[27]将SAC算法应用于离散推荐场景;Rossiiev等^[28]系统综述了基于强化学习的推荐系统,明确指出多模态信息与强化学习框架的深度融合是未

来研究的重要方向。

综上所述,传统推荐系统依赖单一的用户—物品交互矩阵,难以建模多模态语义的互补性^[5]。尽管多模态推荐方法(如 MMGCN^[10]、LATTICE^[29])通过图神经网络或跨模态注意力机制提升了项目表征能力,但仍面临以下关键问题:①数据稀疏性。用户—物品交互天然稀疏,长尾项目缺乏有效交互,导致监督学习易过拟合热门内容;②静态建模局限。现有方法通过固定图结构或注意力权重融合多模态特征(如 VBPR 的视觉特征拼接),无法捕捉用户兴趣的时序依赖^[9];③热门偏差加剧。传统损失函数(如 BPR)通过最大化正负样本的排序差异优化模型,但高频交互项目的主导地位导致长尾内容被系统性忽视,进一步削弱推荐生态的公平性^[30]。为此,本文提出多模态强化学习推荐框架(SAC-Contrastive Multimodal Recommendation, SCMRec),将 SAC 算法与多模态特征融合,并引入对比学习增强策略网络,旨在解决动态推荐场景下的稀疏性与热门偏差问题。本研究的主要创新点为:①建模动态多模态交互。使用 SAC 算法将多媒体推荐建模成 MDP 问题,设计实时更新的用户—物品交互图,突破传统静态图建模的局限性,以捕捉时序演化中的细粒度偏好;②引入对比增强策略网络。在 SAC 的 Actor 网络中嵌入对比损失函数,利用用户行为模式相似性构建正负样本对,解决稀疏数据导致的表征塌缩问题,提升稀疏场景下的判别性。该设计突破了传统强化学习策略网络对显式反馈的依赖,显著降低了过拟合风险;③引入多样性导向的奖励机制。提出长尾加权因子与熵正则化联合优化方案,通过调整热门项目的奖励衰减系数,协同优化推荐准确性、多样性及冷启动能力。

2 SCMRec 模型构建

2.1 问题定义

强化学习使用智能体作为决策机器,其可感知当前环境状态,通过计算生成动作并将其作用于环境中,环境产生相应的奖励信号并发生状态转移,以此获得最大累计奖励期望。将强化学习应用于推荐系统中需要对推荐场景重新建模,即将推荐问题定义为 MDP,使推荐过程成为一个动态决策过程,根据用户当前的状态和历史行为作出最优推荐。MDP 由 5 个关键元素组成,分别为:①状态 s 。在推荐过程中,推荐系统作为智能体执行决策行为,将用户的历史交互信息作为当前环境信息,状态被定义为 $S = \{s | s \in I\}$, I 为用户的历史交互信息;②动作 a 。 a 表示智能体在当前状态下可以采取的行动,在序列推荐中通常指一个或多个物品,即从系统中选出用户大概率会点击的项目。在具体实现中, a 为一个或多个嵌入向量;③状态转移概率 P 。转移概率 $P(s' | (s, a))$ 表示在状态 s 下,智能体采取动作 a 后,用户状态由 s 转移到状态 s' 的概率。在推荐过程中,其

可表现为用户在接收到系统推荐的物品后选择这些物品的概率;④奖励 r 。奖励值 $r(s, a)$ 表示在状态 s 下,智能体采取动作 a 后从环境中获得的即时回报。在推荐过程中,奖励可表现为用户对所推荐物品的点击、购买、评分等行为的量化表示。用户的不同行为被量化为不同的奖励值,从而引导推荐系统的优化方向;⑤折扣因子 γ 。 γ 的取值在 0~1 之间,用于平衡当前奖励与未来奖励的权重。 γ 越小,表明当前奖励在累积奖励中的占比越高,偏向于优化短期收益; γ 越大,表明即时奖励在累积奖励中的占比越高,偏向于优化长期收益。

2.2 模型框架

基于多模态特征与强化学习的 SCMRec 推荐框架结构如图 1 所示。其通过 MDP 建模用户兴趣的动态演化,并利用 SAC 算法的序列决策机制实现用户兴趣的细粒度捕捉与长尾内容的公平推荐。该框架的核心组件包括多模态状态编码、策略网络(Actor)与双 Q 值网络(Critic),三者协同作用,形成闭环的“感知—决策—反馈”优化流程。

2.2.1 多模态状态编码

用户状态 s_t 表示强化学习智能体在时间步 t 感知的环境状态,由多模态嵌入和交互图构成。定义为:

$$s_t = (e_u^{(t)}, e_i^{(t)}, G_{ui}) \quad (1)$$

式中: $e_u^{(t)} \in R^d$ 与 $e_i^{(t)} \in R^d$ 分别表示经过多模态神经网络处理后,用户与物品在 d 维统一隐含空间中的特征嵌入向量; G_{ui} 表示随交互过程实时重构的用户—物品交互图,用于刻画当前环境下节点间的高阶拓扑关系。用户交互图 G_{ui} 实时更新,反映动态交互关系。

多模态用户嵌入和多模态物品嵌入通过多模态神经网络聚合用户交互物品的视觉特征和文本特征得到。表示为:

$$e_u^{(t)}, e_i^{(t)} = MF(f_v, f_t, G_{ui}) \quad (2)$$

式中: $MF(\cdot)$ 表示多模态特征融合函数;视觉特征 f_v 为通过预训练 ResNet-50 提取的物品图像高维语义向量;文本嵌入 f_t 由预训练 BERT 模型对商品描述文本进行深度编码生成。该函数通过在 G_{ui} 上执行图消息传递,实现了异构模态信息与图结构的深度耦合。

2.2.2 策略网络

策略网络 $\pi_\theta(a|s_t)$ 负责根据当前提供的状态 s_t 生成当前最佳的推荐动作。为了在连续动作空间内实现有效的策略探索并确保训练的可微性,本文采用高斯策略建模与重参数化采样技术。网络首先输出动作分布的均值 μ 与标准差 σ ,并据此执行动作采样。公式为:

$$\mu, \sigma = \pi_\theta(s_t), a_t \sim N(\mu, \sigma) \quad (3)$$

式中: π_θ 表示参数为 θ 的深度策略网络; s_t 为 t 时刻输入的综合状态向量; μ 和 σ 分别为网络生成的动作分布均值与标准差; a_t 表示通过重参数化机制采样得到的连续动作向量,其服从均值为 μ 、标准差为 σ 的正态分布 N 。

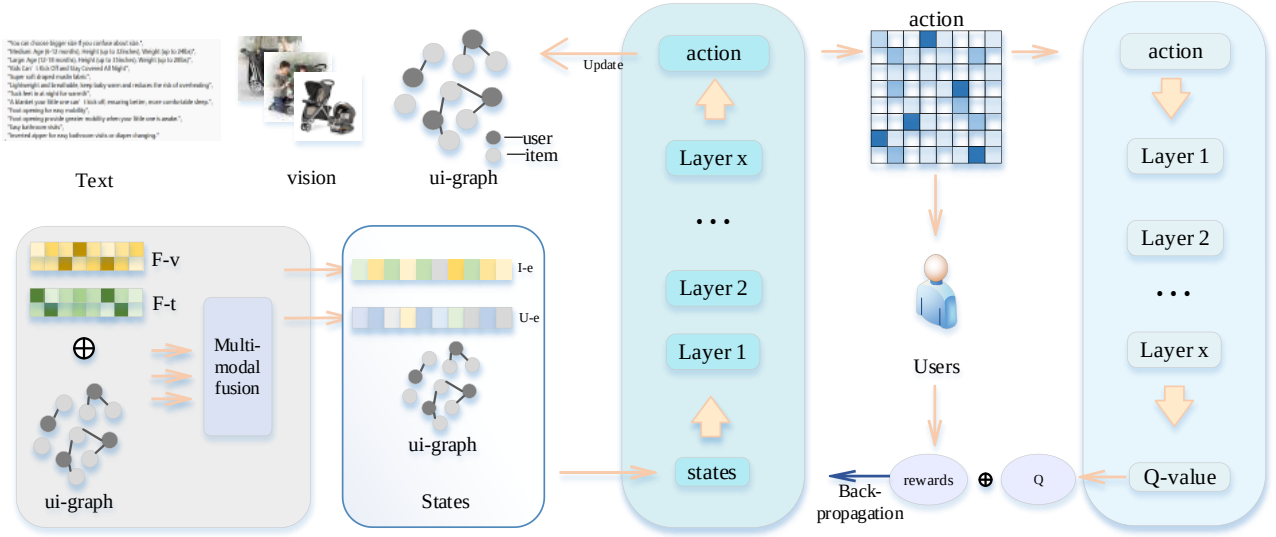


Fig. 1 SCMRec frame structure

图1 SCMRec框架结构

随后,网络通过Top-K动作选择机制从概率分布中筛选出关联度最高的物品,形成最终的推荐列表。表示为:

$$a_i = \text{TopK} \left(\text{Softmax} \left(N(\mu, \sigma) \right) \right) \quad (4)$$

式中: $N(\mu, \sigma)$ 表示前述步骤确立的高斯动作分布; $\text{Softmax}(\cdot)$ 为归一化函数,旨在将分布映射为各个候选物品项的概率得分; $\text{Top-K}(\cdot)$ 算子则依据得分高低筛选出前 K 个目标物品。通过该过程,模型实现了从高维连续意图空间向离散物品推荐序列的决策转换。

2.2.3 双Q值网络

单一Q网络容易因价值过估计而导致策略收敛至局部最优解。为解决这一问题,本文采用双Q价值网络架构,通过悲观估计策略来抑制偏差。系统维护两个独立评估分支 $Q_{\varnothing_1}$ 与 $Q_{\varnothing_2}$, 并取其输出的极小值作为目标价值参考点。计算公式为:

$$Q_{\text{target}} = r_t + \gamma \min_{i=1,2} Q_{\varnothing_i}(s_{t+1}, a_{t+1}) \quad (5)$$

式中: r_t 为执行当前推荐动作后获得的即时奖励; γ 为折扣因子,用于调节未来收益在累积奖励中的权重; s_{t+1} 与 a_{t+1} 分别表示下一时刻的状态与动作。为不断优化价值网络参数并提高评估精准度,系统利用上述目标值作为监督信号来指导网络学习。计算当前网络预测值与目标值之间的均方误差,定义损失函数为:

$$L_Q = \sum_{i=1,2} E_{(s,a,r,s')} \left[\left(Q_{\varnothing_i}(s,a) - Q_{\text{target}} \right)^2 \right] \quad (6)$$

式中: L_Q 表示价值网络的总损失函数; $E_{(s,a,r,s')}$ 表示对经验回放缓冲区中随机采样的状态、动作、奖励、下一状态四元组数据轨迹取期望; $Q_{\varnothing_i}(s,a)$ 表示当前在线价值网络对给定状态—动作对生成的估值。通过最小化该损失函数,两个Q网络分支的参数 $\varnothing_i$ 得以不断修正,从而为策略网络的决策优化提供稳定且可靠的价值评价。

2.3 对比增强策略网络

在动态推荐场景中,稀疏的用户交互数据容易导致策略网络学习到的动作分布缺乏判别性,从而引发表征退化问题。为此,本文提出一种对比增强策略网络,通过引入对比学习机制来约束用户和项目的潜在表示空间,提升策略网络在稀疏数据下的鲁棒性。本文采用批次内对比机制,正样本对由同一用户在不同增强视图下的嵌入表示构成,通过对同一输入施加不同的随机跨模态映射来模拟相似的行为模式;负样本对则由同一训练批次内随机采样的其他用户嵌入构成,利用矩阵乘法隐式构造用户间的互负关联。对比损失旨在最大化正样本对的相似性,同时最小化负样本对的相似性。表示为:

$$L_{CL} = -\log \frac{\exp \left(\frac{s(z_u, z_p)}{\tau} \right)}{\sum_n \exp \left(\frac{s(z_u, z_n)}{\tau} \right)} \quad (7)$$

式中: z_u 为用户嵌入; z_p 和 z_n 分别为正、负样本嵌入; $s(\cdot)$ 为相似度计算函数; τ 为温度系数,用于控制样本分布的平滑度。

2.4 多样性优化

传统推荐系统因过度依赖用户点击率,往往陷入热门偏差,导致长尾项目曝光不足。为此,本文提出一种多样性导向的奖励机制,通过长尾加权因子与熵正则化的联合优化,平衡即时收益与长期生态健康。

2.4.1 奖励设置

针对多模态推荐中普遍存在的热偏差问题,本文引入长尾加权因子,利用物品的历史交互频次来动态调节即时奖励的强度,其目标为抑制热门物品在反馈信号中的主导地位,从而引导推荐策略向具有潜在价值的长尾物品偏移。系统在计算即时奖励 R_t 时,通过引入与流行度相关的

衰减系数,实现对不同热度物品奖励权重的差异化分配。计算公式为:

$$R_i = \text{cosine}(e_u, e_i) \times \frac{1}{\sqrt{\text{Popularity}(i) + 1}} \quad (8)$$

式中: $\text{cosine}(e_u, e_i)$ 为用户嵌入 e_u 与物品嵌入 e_i 的余弦相似度,用于衡量推荐动作的实时匹配精度; $\text{Popularity}(i)$ 为物品 i 在历史交互数据中的被点击或购买频次。对于热门物品($\text{Popularity}(i) \gg 1$),其对应的衰减系数较小,奖励值被显著压缩;而对于长尾物品($\text{Popularity}(i) \approx 0$),其权重系数接近1,确保了其奖励信号不会被热门物品的高频反馈所稀释,从而为长尾物品保留了充分的推荐动力。

为进一步从策略优化层面平衡探索与利用,本文在 SAC 框架中引入熵正则化目标,通过在奖励机制中融入策略熵项来防止智能体过早陷入局部最优。策略损失函数的定义为:

$$L_{\text{policy}} = -\left(Q(s_t, a_t) - \alpha \log \pi(a_t | s_t)\right) \quad (9)$$

式中: α 为自适应温度系数,用于动态调节熵项在损失计算中的权重;熵项 $\log \pi(a_t | s_t)$ 的引入促使策略分布趋于平坦,强制模型尝试更多样化的推荐动作,有效避免了推荐决策因过度拟合热门项目而产生的信息茧房效应。长尾加权奖励与策略熵的深度耦合在保证预测准确性的同时显著提升了推荐结果的公平性与多样性。

2.4.2 联合优化

为使模型在动态推荐过程中兼顾推荐精准性、分布多样性以及在数据稀疏场景下的表征鲁棒性,系统需构建一个能够平衡多个优化目标的联合损失函数。该函数通过标量化加权的方式,对旨在最大化累积奖励与策略熵的策略损失、提升特征判别力的辅助对比损失以及减小价值估计偏差的评价网络损失进行整合。基于此,系统总损失函数 L_{total} 的计算公式为:

$$L_{\text{total}} = L_{\text{policy}} + L_{\text{CL}} + L_Q \quad (10)$$

式中: L_{total} 为全局优化目标; L_{policy} 表示结合策略熵正则化的 Actor 损失,用于引导智能体在策略空间内执行广泛探索; L_{CL} 表示基于自监督学习的对比损失,旨在利用结构一致性约束抑制表征塌缩; L_Q 为 Critic 网络的均方差损失,负责量化状态—动作对的价值评估误差。通过最小化该联合函数,模型实现了在多维性能指标上的协同收敛。

在非平稳的动态交互环境下,为确保贝尔曼目标值的平稳演变并抑制梯度震荡,系统引入指数移动平均驱动的软更新机制进行参数同步。该机制的核心逻辑为使目标网络参数以极小的步长追踪在线网络变化,从而在数学上构建一个稳定的函数近似锚点。参数更新公式为:

$$\phi' \leftarrow \tau \phi + (1 - \tau) \phi' \quad (11)$$

式中: ϕ 为当前在线决策网络的权重参数; ϕ' 为对应的目标网络参数; τ 为平滑系数。在该更新范式下,通过限制 τ 的取值范围($0 < \tau \ll 1$),可以确保目标网络参数在时

间轴上实现连续且缓慢的更迭。这种设计有效降低了由状态转移带来的非平稳性影响,保障了强化学习框架在处理复杂多模态特征时的数值稳定性。

2.5 训练流程

基于 SAC 的推荐策略训练算法伪代码为:

输入:用户交互数据 G_{ui} 、多模态特征 f_u, f_i ;

输出:策略网络参数 θ 、Q 网络参数 ϕ ;

初始化:策略网络 π_θ 、Q 网络 Q_ϕ 、目标网络 $Q_{\text{target}} \leftarrow Q_\phi$;

1. FOR 每个周期 DO:

2. 初始化状态 $s_0 = (e_u^{(0)}, e_i^{(0)}, G_{ui}^{(0)})$;

3. FOR 每个时间步 DO:

4. 生成动作: $a_t, \text{top_k_idx} \leftarrow \pi_\theta(s_t)$, 公式(3);

5. 执行推荐,计算奖励 $r_t = \text{相似度奖励} \times \text{长尾权重}$, 公式(8);

6. 存储经验: $(s_t, a_t, r_t, s_{t+1}) \rightarrow \text{回放缓冲区}$;

7. 从缓冲区采样批量数据;

7. 更新 ϕ 以最小化 $\text{MSE}(Q_{\text{target}}(s_t, a_t), Q_{\text{target}})$, 公式(5);

8. 更新 θ 以最小化 L_{total} , 公式(10);

9. $\phi' \leftarrow \tau \phi + (1 - \tau) \phi'$, 公式(11);

10. END

11.END

3 验证实验

3.1 实验环境与参数设置

所有实验均在高性能服务器环境中完成。硬件采用主频为 2.10 GHz 且具备 12 核心的 Intel(R) Xeon(R) Platinum 8352V 中央处理器,辅以 90 GB 运行内存以及 30 GB 存储空间,同时配备 1 块显存容量为 32 GB 的虚拟化 GPU 以支撑大规模运算需求。在软件配置方面,基于 Ubuntu 20.04 操作系统安装 Python 3.8 开发环境,并利用 PyTorch 1.11.0 深度学习框架结合 CUDA 11.3 运算平台共同实现了算法的训练与推理加速。

SCMRec 选择 Adam^[31]作为优化器来优化模型,学习率搜索范围为 $\{4.5\text{e-}4, 5\text{e-}4, 5.5\text{e-}4, 6\text{e-}4\}$ 。为进行公平比较,将默认批量大小 batch_size 设置为 1 024,嵌入维度设置为 64。对于所有基于图的方法,协作图传播层的数量 L 设置为 2。采用早期停止机制,即当验证集上的 R@20 连续 10 个时期不增加时,训练将停止。

3.2 数据集

使用 3 个公开可用的多模态推荐数据集,分别为 Amazon-Baby、Amazon-Sports 和 Allrecipes。其中,Amazon-Baby 和 Amazon-Sports 包含由 Sentence-BERT 模型基于商品标题、描述、品牌和类别信息生成的文本特征嵌入^[32];Allrecipes 数据集来源于全球最大的食品类社交平台之一,涵盖 27 种类别,共计 52 821 道食谱,每个食谱的图片和食材信息分别被视为其视觉特征和文本特征^[33]。以上数据集

覆盖了商品、食品等多样化场景,能全面验证模型的泛化能力。表1为3个数据集的基本信息。

Table 1 Basic information of datasets

表1 数据集基本信息

数据集	用户数	物品数	交互数	稀疏度/%
Amazon-Baby	19 445	7 050	139 110	99.89
Amazon-Sports	35 598	18 357	296 337	99.95
Allrecipes	19 805	10 067	58 922	99.97

3.3 基线方法与评价指标

为验证SCMRec的有效性,将其与传统推荐和多媒体推荐两类方法进行比较。传统推荐方法包括:①BPR^[30]。该方法利用贝叶斯个性化排名损失来提高基于矩阵分解的性能;②LightGCN^[34]。该方法通过剔除非线性变换与权重矩阵来简化图卷积架构,专注于利用邻域聚合机制捕获图结构中本质的高阶协同过滤信号,从而提升模型的泛化能力与训练效率。多媒体推荐方法包括:①VBPR^[9]。该方法将预训练的视觉特征作为辅助信息引入矩阵分解框架,以增强对项目属性的表示深度;本研究遵循通用实验规范,利用文本模态对其进行了语义扩展,以确保评估的全面性;②MMGCN^[10]。该方法利用图卷积网络为每种模态分别构建特定的特征图,通过跨模态的消息传递捕捉模态特有的高阶关联,并最终通过多模态特征融合生成预测结果;③LATTICE^[29]。该方法通过挖掘项目间的潜在语义关系来构建额外的项目关联图,通过结构化的方式捕捉隐含在多模态特征空间中的高阶语义相关信号;④SLMRec^[12]。该方法首次引入自监督学习范式处理多模态推荐,通过设计特征噪声扰动与多模态模式发现等辅助任务,增强了模型在数据稀疏场景下的表征鲁棒性;⑤MMSL^[13]。该方法结合自监督学习与对比学习理论,通过设计模态感知的交互结构学习与跨模态对比任务,有效强化了异构特征空间之间的一致性约束;⑥DiffMM^[17]。该方法

利用多模态扩散模型生成模态感知的用户-项目图,并结合跨模态对比学习,有效对齐多模态特征与用户交互。

为评估各方法的推荐准确性,采用Recall@K (R@K)、Precision@K (P@K)和归一化折损累计增益NDCG@K (N@K)这3种广泛使用的指标。参考文献[10,35]的设置,采用全排序评估协议,并设定 $K=20$,最终计算测试集上所有用户的平均性能。

3.4 实验结果分析

3.4.1 不同方法推荐性能比较

表2为不同方法的推荐性能比较,采用粗体突出显示SCMRec的性能,性能最佳的基线方法以下划线展示。可以看出,SCMRec在3个数据集上的推荐性能均优于其他基线方法。这得益于强化学习框架的动态建模能力与对比学习增强的表示学习机制。通过MDP,SCMRec能捕捉用户与项目的时序交互动态,并结合长期奖励机制优化探索与利用的平衡。此外,在策略网络中引入对比学习进一步强化了用户和项目的潜在表示,尤其在稀疏数据场景下表现突出。传统推荐方法(如BPR、LightGCN)通过静态交互建模实现推荐,但难以捕捉用户行为的时序动态。相比之下,SCMRec通过SAC算法最大化长期累积奖励,动态调整推荐策略,从而更自然地融合多模态特征与用户行为序列。例如,相较于静态建模的VBPR和MMGCN,SCMRec在Sports数据集上的Recall@20显著提升。多媒体推荐方法(如MMSL、LATTICE)通过跨模态对抗生成或预定义关联增强表示,但可能忽略用户行为的动态性。而SCMRec将对比学习嵌入策略网络,该机制可直接约束策略网络生成的动作分布,拉近相似用户的嵌入表示,推远无关用户,从而增强语义一致性并缓解稀疏性问题。例如,在长尾特征显著的Allrecipes数据集上,SCMRec表现出更强的鲁棒性。

Table 2 Comparison of recommended performance of different methods

表2 不同方法推荐性能比较

模型	Amazon-Baby			Amazon-Sports			Allrecipes		
	R@20	N@20	P@20	R@20	N@20	P@20	R@20	N@20	P@20
BPR	0.044 0	0.020 0	0.002 4	0.043 0	0.020 2	0.002 3	0.013 7	0.005 3	0.000 7
lightGCN	0.069 8	0.031 9	0.003 7	0.078 2	0.036 9	0.004 2	0.021 2	0.007 6	0.001 0
VBPR	0.048 6	0.021 3	0.002 6	0.058 2	0.026 5	0.003 1	0.015 9	0.005 6	0.000 8
MMGCN	0.061 5	0.026 4	0.003 1	0.063 8	0.027 9	0.003 4	0.026 1	0.010 1	0.001 3
LATTICE	0.084 3	0.036 8	0.004 2	0.091 5	0.042 4	0.004 8	0.026 8	0.010 3	0.0014
SLMRec	0.076 5	0.032 5	0.004 3	0.082 9	0.037 6	0.004 3	0.025 2	0.009 9	0.001 3
MMSL	0.093 6	<u>0.040 5</u>	0.004 8	0.098 3	0.045 5	0.005 2	0.034 1	0.013 2	0.001 7
DiffMM	<u>0.094 7</u>	0.040 1	<u>0.004 9</u>	<u>0.101 7</u>	<u>0.045 8</u>	<u>0.005 4</u>	<u>0.035 2</u>	<u>0.013 5</u>	<u>0.001 7</u>
SCMRec	0.098 4	0.045 4	0.005 4	0.103 7	0.048 5	0.005 5	0.038 0	0.014 0	0.001 9

3.4.2 消融实验

通过消融实验验证各组件对SCMRec模型性能的情况,结果见表3。可以看出,移除动态图更新模块后,w/o Dynamic-G模型在Amazon-Baby、Amazon-Sports、Allrecipes数据集上的Recall@20和NDCG@20均大幅下降,表明动态更新用户-项目交互图对于捕捉用户行为的时序变

化至关重要。固定交互图会导致模型无法适应兴趣演变,尤其在稀疏场景下(如Allrecipes数据集上)性能损失显著。将SAC替换为其他强化学习算法(DQN)后,w/o SAC模型性能全面下降,这是由于SAC通过最大化熵正则化目标,在探索(推荐多样性)与利用(推荐准确性)之间实现更优平衡,尤其在复杂动态场景中(如Amazon-Sports数据集

上的高交互密度数据)表现更稳定。禁用策略网络中的对比学习模块后,w/o CL模型在稀疏数据集 Allrecipes 上的 Recall@20 和 NDCG@20 均有所下降,表明对比学习通过拉近相似用户嵌入、推远无关用户提升了稀疏场景下的判别性。在密集数据集(如 Amazon-Sports)中,NDCG@20 因预定义关联的干扰出现异常波动,进一步验证了对比学习需结合动态交互建模以避免静态偏置。完整模型 SCMRec 通过动态图更新捕捉用户行为时序变化,通过 SAC 算法优化长期奖励,通过对比学习增强稀疏数据表示,三者协同作用使 SCMRec 在不同类型数据集上均达到最优性能,尤其在长尾场景(Allrecipes 数据集)中,动态交互与对比学习的联合设计缓解了数据稀疏性,提高了模型鲁棒性。

Table 3 Ablation experiment results

表 3 消融实验结果

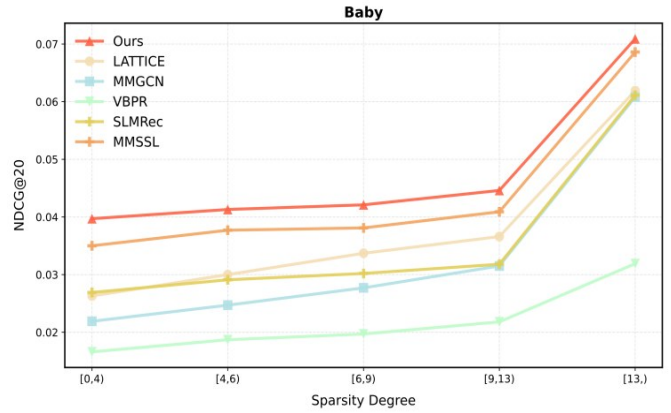
消融 模型	Amazon-Baby		Amazon-Sports		Allrecipes	
	R@20	N@20	R@20	N@20	R@20	N@20
w/o Dynamic -G	0.091 8	0.040 5	0.096 6	0.044 7	0.033 5	0.012 4
w/o SAC	0.093 3	0.040 7	0.098 4	0.045 0	0.034 2	0.013 2
w/o CL	0.096 0	0.043 2	0.101 7	0.047 3	0.036 2	0.013 5
SCMRec	0.098 4	0.045 4	0.103 7	0.048 5	0.038 0	0.014 0

3.4.3 数据稀疏性评估

为验证 SCMRec 模型解决数据稀疏性问题的有效性,以 NDCG@20 作为衡量标准,在不同用户交互频率下评估其推荐准确性,结果见图 2。可以看出,各模型的 NDCG@20 在 Amazon-Baby 数据集上随交互次数增加呈稳步上升趋势,而在 Allrecipes 数据集上则表现出波动的增长趋势。这说明交互信息的丰富程度整体上有利于模型学习,但在极端稀疏环境下仍存在性能波动。Allrecipes 作为极稀疏数据集,其多模态信息较为复杂,在交互密度处于中低水平时,新增交互可能引入一定的拓扑噪声,导致静态建模的模型出现表征波动。相比之下,SCMRec 在该区间内表现得更为稳健,曲线维持了上升或持平状态。综合来看,SCMRec 在两个数据集的所有稀疏度分组下均优于基线模型,在低频交互场景中展现出更强的稳健性。这种在稀疏场景下的鲁棒性归因于 SAC 框架的动态决策能力与对比增强机制的有机结合。通过在动态重构的图结构上进行自监督学习,模型能更有效地在低交互环境下提取本质的偏好特征,从而有效缓解数据稀疏性问题。

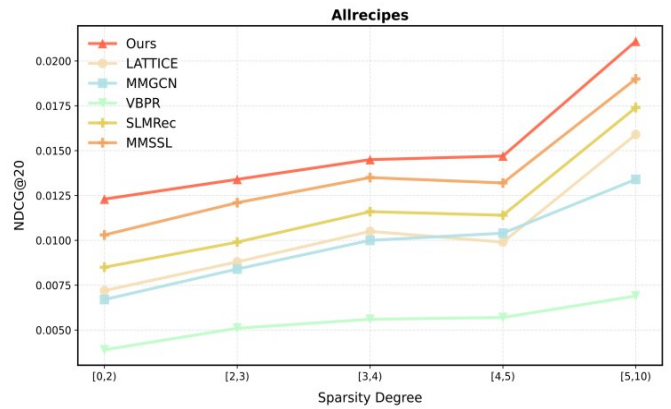
4 结语

本文通过结合强化学习 SAC 算法与对比学习策略,提出一种新型多媒体推荐模型 SCMRec,为稀疏场景下的多媒体推荐提供了高效解决方案。实验结果表明,该模型在极端稀疏场景(如 Allrecipes 数据集的 99.97% 稀疏度)下仍能保持较强的鲁棒性,在实际应用中将有广阔的前景和巨大的潜力。然而,本研究仍存在一些局限性,例如 SCMRec



(a) Amazon-Baby dataset

(a) Amazon-Baby 数据集



(b) Allrecipes dataset

(b) Allrecipes 数据集

Fig. 2 NDCG@20 of each method under different sparsity levels

图 2 不同稀疏度下各方法的 NDCG@20

模型在训练过程中对计算资源的要求较高,尤其是在构建和动态更新用户-项目交互图时的时间复杂度过高,可能限制其在超大规模实时推荐系统中的应用。后续可以探索更高效的图结构更新策略,如引入增量学习机制或简化网络结构,以降低计算开销并提升模型的实用性。

参考文献:

- [1] WU S, SUN F, ZHANG W, et al. Graph neural networks in recommender systems: a survey[J]. ACM Computing Surveys, 2022, 55(5): 1-37.
- [2] XU J, CHEN Z, YANG S, et al. A survey on multimodal recommender systems: recent advances and future directions[DB/OL]. <https://arxiv.org/abs/2502.15711>.
- [3] LI H, LI C, FENG K, et al. Robust knowledge adaptation for dynamic graph neural networks[J]. IEEE Transactions on Knowledge and Data Engineering, 2024, 36(11): 6920-6933.
- [4] YU J, YIN H, XIA X, et al. Self-supervised learning for recommender systems: a survey[J]. IEEE Transactions on Knowledge and Data Engineering, 2023, 36(1): 335-355.
- [5] NAJMANI K, BENLAHMAR E H, SAEL N, et al. Collaborative filtering approach: a review of recent research[C]//International Conference on Advanced Intelligent Systems for Sustainable Development, 2020: 151-163.
- [6] MALITESTA D, CORNACCHIA G, POMO C, et al. Formalizing multi-

- media recommendation through multimodal deep learning[J]. *ACM Transactions on Recommender Systems*, 2025, 3(3): 1–33.
- [7] Kaelbling L P, Littman M L, Moore A W. Reinforcement learning: a survey[J]. *Journal of Artificial Intelligence Research*, 1996, 4: 237–285.
- [8] Shani G, Heckerman D, Braffman R I. An MDP-based recommender system[J]. *Journal of Machine Learning Research*, 2005, 6(9): 1265–1295.
- [9] He R, McAuley J. VBPR: visual Bayesian personalized ranking from implicit feedback[C]//*Proceedings of the AAAI Conference on Artificial Intelligence*, 2016: 144–150.
- [10] Wei Y, Wang X, Nie L, et al. MMGCN: multi-modal graph convolution network for personalized recommendation of micro-video[C]//*Proceedings of the 27th ACM International Conference on Multimedia*, 2019: 1437–1445.
- [11] Wei Y, Wang X, Nie L, et al. Graph-refined convolutional network for multimedia recommendation with implicit feedback[C]//*Proceedings of the 28th ACM International Conference on Multimedia*, 2020: 3541–3549.
- [12] Tao Z, Liu X, Xia Y, et al. Self-supervised learning for multimedia recommendation[J]. *IEEE Transactions on Multimedia*, 2022, 25: 5107–5116.
- [13] Wei W, Huang C, Xia L, et al. Multi-modal self-supervised learning for recommendation[C]//*Proceedings of the ACM Web Conference*, 2023: 790–800.
- [14] Chen J, Dong H, Wang X, et al. Bias and debias in recommender system: a survey and future directions[J]. *ACM Transactions on Information Systems*, 2023, 41(3): 1–39.
- [15] Hu H, Wang X, Zhang Y, et al. A comprehensive survey on contrastive learning[J]. *Neurocomputing*, 2024, 610(c): 128645.
- [16] Le-Khac P H, Healy G, Smeaton A F. Contrastive representation learning: a framework and review[J]. *IEEE Access*, 2020, 8: 193907–193934.
- [17] Jiang Y, Xia L, Wei W, et al. Diffmm: multi-modal diffusion model for recommendation[C]//*Proceedings of the 32nd ACM International Conference on Multimedia*, 2024: 7591–7599.
- [18] Watkins C J C H, Dayan P. Q-learning[J]. *Machine Learning*, 1992, 8: 279–292.
- [19] Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning[J]. *Nature*, 2015, 518(7540): 529–533.
- [20] Lillicrap T P, Hunt J J, Pritzel A, et al. Continuous control with deep reinforcement learning [DB/OL]. <https://arxiv.org/abs/1509.02971>.
- [21] Zhao D, Wang H, Shao K, et al. Deep reinforcement learning with experience replay based on SARSA[C]//2016 IEEE Symposium Series on Computational Intelligence, 2016: 1–6.
- [22] Wei Z, Xu J, Lan Y, et al. Reinforcement learning to rank with Markov decision process[C]//*Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2017: 945–948.
- [23] Zheng G, Zhang F, Zheng Z, et al. DRN: a deep reinforcement learning framework for news recommendation[C]//*Proceedings of the 2018 World Wide Web Conference*, 2018: 167–176.
- [24] Haarnoja T, Zhou A, Hartikainen K, et al. Soft actor-critic algorithms and applications[DB/OL]. <https://arxiv.org/abs/1812.05905>.
- [25] Chen H, Dai X, Cai H, et al. Large-scale interactive recommendation with tree-structured policy gradient[C]//*Proceedings of the AAAI Conference on Artificial Intelligence*, 2019: 3312–3320.
- [26] Liu L, Guan Y, Wang Z, et al. An interactive food recommendation system using reinforcement learning[J]. *Expert Systems with Applications*, 2024, 254: 124313.
- [27] Luo J, Li F, Jiao J. A dynamic multiobjective recommendation method based on soft actor-critic with discrete actions[J]. *Journal of King Saud University Computer and Information Sciences*, 2025, 37(1): 1–17.
- [28] Rossiev O D, Shapovalova N N, Rybalchenko O H, et al. A comprehensive survey on reinforcement learning-based recommender systems: state-of-the-art, challenges, and future perspectives[C]//*CEUR Workshop Proceedings*, 2025: 428–440.
- [29] Zhang J, Zhu Y, Liu Q, et al. Mining latent structures for multimedia recommendation[C]//*Proceedings of the 29th ACM International Conference on Multimedia*, 2021: 3872–3880.
- [30] Rendle S, Freudenthaler C, Gantner Z, et al. BPR: Bayesian personalized ranking from implicit feedback[DB/OL]. <https://arxiv.org/abs/1205.2618>.
- [31] McAuley J, Targett C, Shi Q, et al. Image-based recommendations on styles and substitutes[C]//*Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2015: 43–52.
- [32] Gao X, Feng F, He X, et al. Hierarchical attention network for visually-aware food recommendation[J]. *IEEE Transactions on Multimedia*, 2019, 22(6): 1647–1659.
- [33] Wei Y, Wang X, He X, et al. Hierarchical user intent graph network for multimedia recommendation[J]. *IEEE Transactions on Multimedia*, 2021, 24: 2701–2712.
- [34] He X, 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*, 2020: 639–648.
- [35] Kingma D P, Ba J. Adam: a method for stochastic optimization[DB/OL]. <https://arxiv.org/abs/1412.6980>.

(责任编辑:尹晨茹)