特征掩码与对比学习融合多维度去过度相关的序列推荐

钱忠胜*,刘金平,李玉龙,范赋宇,陈 超
(江西财经大学计算机与人工智能学院,江西南昌 330013)

摘 要: 近年来,基于自注意力机制的序列推荐模型在用户行为建模中展现出了显著的效果,尤其是在处理长程依赖关系方面表现突出。然而,此类模型(如Transformer类方法)在深层编码过程中,高阶表示会因多层聚合而逐渐趋同,个性化信号被削弱,从而产生过度平滑问题,且该问题在此类模型中往往被忽视;与此同时,特征维度间的高度相关会带来冗余与噪声传播,削弱模型对重要特征的辨识能力,从而进一步限制模型的泛化能力。为此,本文提出特征掩码与对比学习融合多维度去过度相关的序列推荐模型(feature Masking and Contrastive learning integrating Multi-Dimensional decorrelation in Sequential Recommendation,MCMD-SR)。首先,设计自注意力感知的特征掩码机制,依据自注意力得分衡量各维度贡献,对低贡献且易导致表示相似化的特征进行针对性遮蔽,并提出随层数递减的对数掩码率衰减策略,使浅层施加强扰动以打破局部高相似特征,深层保持适度扰动以持续抑制过度聚合。进一步地,在掩码后的最终层表示与原始浅层表示之间构建对比学习任务,通过拉近同一序列的正样本对、推远不同序列或不同特征的负样本对,强化差异化与个性化语义,提升嵌入空间的区分能力。其次,提出多维度自适应去过度相关模块,在注意力掩码后的特征矩阵上分别从列间与层间计算皮尔逊相关系数(Pearson Correlation Coefficient,PCC),并依据相关强度自适应分配惩罚权重,在保持总体正则化强度可控的同时抑制冗余维度与冗余层间依赖,从局部(列间)与全局(层间)双视角降低特征冗余,提升关键特征辨识度。最后,将自注意力掩码机制、对比学习模块与多维度自适应去过度相关模块的损失进行多任务联合优化,使三类约束相互补充,稳定训练并提升嵌入质量与模型泛化性。在Beauty、Yelp、Last-FM和ML-1M四个公开数据集上,本文模型与11个经典及最新的序列推荐模型进行了对比。实验结果表明,在命中率(Hit Ratio, HR)和归一化折损累积增益(Normalized Discounted Cumulative Gain,NDCG)两个指标上,所提模型MCMD-SR相对已有最优基线模型分别平均最少提升2.13%和1.67%,验证了本文模型在推荐性能上的有效性。此外,本文还通过消融实验和参数敏感性实验分析,验证了各模块的必要性及其协同有效性,进一步阐明了模型具有良好的泛化能力。

关键词: 序列推荐;特征掩码;对比学习;去过度相关;自注意力机制;相关性度量

基金项目: 国家自然科学基金(No.62262025);赣鄱俊才支持计划-主要学科学术和技术带头人培养项目-领军人才(学术类)(No.20243BCE51024)

中图分类号: TP399 **文献标识码:** A **文章编号:** 0372-2112(2026)02-0875-24

电子学报URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20250551

Feature Masking and Contrastive Learning Integrating Multi-Dimensional Decorrelation in Sequential Recommendation

QIAN Zhongsheng*, LIU Jinping, LI Yulong, FAN Fuyu, CHEN Chao
(*School of Computing and Artificial Intelligence*, *Jiangxi University of Finance and Economics*, *Nanchang*, *Jiangxi* 330013, *China*)

Abstract: Recently, self-attention-based sequential recommendation models have demonstrated remarkable effectiveness in user behavior modeling. However, these models tend to suffer from an over-smoothing problem during deep encoding. Repeated aggregation across multi layers makes high-order representations become increasingly similar, which gradually weakens personalized signals. Meanwhile, the high correlation among feature dimensions introduces redundancy and noise propagation, which weakens the model's ability to identify important features and consequently limits its generalization capability. To address these challenges, this work proposes feature masking and contrastive learning integrating multi-dimensional decorrelation in sequential recommendation (MCMD-SR), a feature masking and contrastive learning model integrating multi-dimensional decorrelation in sequential recommendation. Firstly, we design a feature masking mechanism based on self-attention. This mechanism measures the contribution of each feature dimension with attention scores. It then selectively masks features with low-contribution which are prone to inducing representation homogenization. In addition,

收稿日期:2025-06-24;录用日期:2025-12-26;责任编辑:覃怀银
*通信作者:钱忠胜

we also introduce a logarithmic mask-rate decay strategy across layers. This strategy applies stronger perturbations in shallow layers to break high-similarity features locally. In deeper layers, it maintains moderate perturbations to continuously suppress excessive aggregation. Furthermore, a contrastive learning task is constructed between the masked final-layer representations and the original shallow-layer representations. The proposed method pulls together positive pairs from the same sequence and pushes apart negative pairs from different sequences or feature dimensions. The proposed method reinforces discriminative and personalized semantics. Thereby it improves the separability of the embedding space. Secondly, we propose a multi-dimensional adaptive decorrelation module. Based on the attention-masked feature matrix, Pearson correlation coefficients (PCC) are computed from both column-wise and layer-wise perspectives. Penalty weights are adaptively assigned according to the correlation strength. This suppresses redundant dimensions and inter-layer dependencies. Meanwhile, it keeps the overall regularization strength controllable. This dual-view decorrelation strategy reduces feature redundancy from both local (column-wise) and global (layer-wise) perspectives resulting in improving the identification of key features. Finally, the self-attention masking mechanism, the contrastive learning module, and the multi-dimensional adaptive decorrelation module are jointly optimized in a multi-task learning framework. These complementary constraints stabilize training and improve embedding quality as well as model generalization. Extensive experiments are conducted on 4 public datasets, where the proposed method is compared with 11 classical and state-of-the-art sequential recommendation models. Experimental results show that MCMD-SR achieves average improvements of 2.13% and 1.67% over the strongest baseline in terms of hit ratio (HR) and normalized discounted cumulative gain (NDCG), respectively. In addition, ablation studies and parameter sensitivity analysis further verify the necessity of each module and their synergistic effectiveness, thereby further clarifying the strong generalization capability of our model.

Keywords: sequential recommendation; feature masking; contrastive learning; decorrelation; self-attention mechanism; correlation metrics

Foundation Item(s): National Natural Science Foundation of China (No.62262025); Jiangxi Poyang Support Program for Talents of China-Major Discipline Academic and Technical Leaders Training Program of Jiangxi Province-Leading Talent (Academic) (No.20243BCE51024)

0 引言

序列推荐(Sequential Recommendation,SR)因其在时序理解以及用户行为建模等方面展现出的显著优势而被广泛关注。相较于传统推荐方法,SR通过分析用户的时序信息,可更准确地捕捉用户的兴趣演变和动态变化,从而为用户提供更准确的推荐。

SR主要关注如何从历史行为序列中挖掘用户的潜在偏好,从而预测用户可能感兴趣的内容。近来,基于自注意力机制的SR方法(如Transformer)往往通过对用户历史行为序列进行建模,以捕捉用户兴趣的演化轨迹,在SR任务中取得显著进展。然而,这类方法在建模过程中需对用户与项目的交互序列进行高阶表示和学习,使得对不同的行为序列进行编码时特征逐渐趋于相似,导致个性化信息的丢失,进而引起过度平滑(Over-smoothing)问题。Shin等人[1]首次在SR场景下系统性地证明了自注意力机制会不断削弱高频行为模式并导致表示趋同,从而引发过度平滑现象,进而使模型难以捕获用户的兴趣波动。此外,Fan等人[2]从奇异值谱衰减的角度揭示了序列推荐模型在深层训练时表示逐渐相似化的现象。与此同时,Zhou等人[3]和Du等人[4]从频域视角指出现有SR模型普遍存在低频成分占优、高频行为表达不足的问题,导致模型难以建模细粒度、突发性的用户偏好变化。现有基于注意力机制的SR方法如CaDiRec(Context-aware Diffusion-based contrastive learning for sequential Recommendation)[5]、MSDCCL(Multi-level Sequence Denoising with Cross-signal Contrastive Learning)[6]、FENRec(Future data utilization with Enduring Negatives for contrastive learning in sequential Recommendation)[7]虽有效地提升了模型的精度,但它们在建模过程中采用高阶表示学习策略,忽略了过度平滑现象对用户真实兴趣偏好表示的潜在影响。可喜的是,图神经网络(Graph Neural Network,GNN)因其强大的信息聚合能力而被用于协同过滤推荐[8],使用GNN来学习用户和项目的嵌入,能学习到更丰富的序列交互信息。但同时GNN会带来过度平滑问题,导致信息不可区分。通过缓解GNN中的过度平滑问题,可提升模型性能[9-13]。故在基于注意力机制的SR中缓解过度平滑问题能避免特征过于相似,使特征更具区分性且使模型训练更稳定。因此,深入探讨并缓解基于注意力机制的SR中过度平滑问题,成为提升推荐性能的一个重要方向。

当前已有方法如CADSR(Context-Aware Diffusion-based Sequential Recommendation)[14]、Query-SeqRec(Query-aware Sequential Recommendation)[15]、AutoSAM

(Automatic SAMpling framework for sequential recommendation)[16]在提升基于注意力机制的SR模型泛化性方面取得了一定进展,然而这些方法普遍忽略了特征维度过度相关性对序列建模带来的不利影响。在基于注意力机制的SR中,特征维度间的高度相关易引入冗余信息,进而引发噪声传播,对序列表示造成干扰,削弱模型对用户偏好的刻画能力。针对该问题,已有研究尝试利用注意力机制进行信息筛选,通过选择性地聚合邻居特征,或特征选择/降维机制压缩冗余维度,或通过跳跃连接保留原始特征表示等方式,有效缓解了冗余干扰,提升了模型的泛化性[17-20]。此外,现有工作对特征维度过度相关性与过度平滑之间的耦合关系关注较少。然而,在深层模型结构中二者相互补充、相互促进。因此,在序列推荐任务中引入特征去相关机制,不仅有助于减少无效或冗余信息的传播,还可与缓解过度平滑策略形成动态互补关系,协同提升模型的泛化能力与推荐准确性。

通过以上分析,我们发现已有基于注意力机制的SR模型存在以下不足:

(1)模型忽视潜在的过度平滑问题,未能从缓解该现象的角度出发进行有效建模,从而限制了表示能力的提升。

(2)已有研究往往忽略特征维度间高度相关性所导致的冗余问题,且未建模该问题与过度平滑之间的潜在耦合关系,这在一定程度上抑制模型的泛化能力。

为应对上述两个方面的问题,本文提出一种特征掩码与对比学习融合多维度去过度相关的序列推荐模型(feature Masking and Contrastive learning integrating Multi-Dimensional decorrelation in Sequential Recommendation,MCMD-SR)。主要工作与贡献如下:

(1)构建自注意力特征掩码结合对比学习模块,提升模型对差异项目的辨识能力以及对潜在个性偏好的建模性能,从而缓解过度平滑问题。

(2)构建多维度自适应去过度相关模块,从局部和全局两个层面建模特征间的相关性,动态过滤冗余信息,提高模型的泛化能力。

(3)在4个公开数据集上展开了对比与消融实验,与当前经典的、主流的序列推荐模型相比,本文模型MCMD-SR在性能上优势较明显,也验证了模型中各组件存在的必要性及其组合的合理性。

1　相关工作

近来,SR已成为推荐领域中的核心任务,主要关注建模用户历史行为序列中的动态偏好。早期方法如GRU4Rec(Gated Recurrent Unit for Recommendation)[21]利用循环神经网络(Recurrent Neural Network,RNN)捕捉行为序列中的时间依赖性,后续如SASRec(Self-Attentive Sequential Recommendation)[22]引入Transformer架构以建模更长范围序列依赖,成为当前主流框架之一。目前,为增强用户行为建模能力,引入多任务学习[23]、GNN[24]、稀疏注意力机制[25]等,显著提升推荐准确性。然而,在基于自注意力机制的SR模型中,随着网络深度增加或特征交互增强,特征表示日益趋同,导致用户个性化信息被削弱,过度平滑问题逐渐引起关注。已有研究尝试从特征掩码、对比学习角度缓解这一问题,同时通过去特征维度过度相关性提升模型的泛化能力。

(1)特征掩码与对比学习缓解过度平滑的方法。通过引入特征掩码机制可增强模型对特征的建模能力,从而提升推荐效果。对比学习作为特征增强的另一种方法,近来成为增强表示学习的一种强有力工具,尤其在个性化建模情形下表现突出。其核心思想是将正样本拉近、负样本推远,从而提升嵌入空间的结构性表达。Zhu等人[26]设计AdaF^{2}M^2(Adaptive Feature Modeling with Feature Mask)框架,通过随机掩码部分特征,减少对重要特征的过度依赖,以增强模型的鲁棒性。Sang等人[27]提出MCL(Masked Contrastive Learning)模型,通过随机掩码策略增强图结构,减少节点对特定邻居敏感性,并增强嵌入健壮性。Xia等人[28]提出TransAct(Transformer-based realtime user Action)推荐方法,通过Transformer编码短期行为序列,结合随机时间窗口掩码策略增强推荐多样性,并与长期兴趣表征融合,提升推荐效果。Zhou等人[29]设计MVCrec(MultiView Contrastive learning for sequential recommendation)序列推荐框架,通过融合ID视图和图视图的双重对比学习,结合多视图注意力机制动态整合两种表示,显著提升了用户偏好建模的区分能力。Qiu等人[30]提出DuoRec推荐方法,通过双塔结构分别建模原始序列与扰动序列,强调在对比学习中保持序列语义一致性,展示更强的抗噪声能力。Peng等人[31]提出TSC(Two-Sided Constraint)模型,在特征嵌入矩阵上随机屏蔽一些特征,对屏蔽后的特征嵌入矩阵进行对比学习,从而保持节点特征的多样性和区分性,提升特征嵌入质量。

可见,现有特征掩码策略多采用随机掩码机制,未能结合注意力分布中的语义信息,难以有效识别并屏蔽对最终表示产生干扰的相似特征,这在一定程度上加剧过度平滑问题。同时,对比学习在增强个性化表示建模方面展现出较强能力,且在缓解过度平滑现象方面具有潜力。基于此,将自注意力掩码机制与对比学习相结合,有望实现更有效的个性化建模,从而

在缓解过度平滑现象的同时提升推荐性能。

(2)去特征维度过度相关性的方法。特征维度过度相关会存在大量冗余,增加噪声传播风险,去特征过度相关可提高模型泛化性。Zhu 等人[32]构建一种去相关正则化方法,动态地选择哪些滤波器需进行去相关操作,避免无用滤波器的干扰。Zeng 等人[33]提出一种新的结构化丢弃方法 CorrDrop(Correlation based Dropout),用于对卷积神经网络(Convolutional Neural Network, CNN)进行正则化,与传统的丢弃方法不同,CorrDrop 根据特征图中的特征相关性丢弃特征单元,以避免过度丢弃或不足丢弃的问题。Jin 等人[34]设计 DeCorr 框架,旨在直接减少特征维度之间的相关性,同时最大化输入与表示之间的互信息,从而帮助深度学习模型更好地编码有用信息。Lin 等人[35]构建 ISA(Iterative Sparse Attention)模型,通过稀疏注意力机制来减少特征之间的相关性,提升推荐系统的性能。Fei 等人[36]提出 EVI(Entire-space Variational Information)推荐方法,利用变分信息挖掘的策略优化特征表示,减少特征相关性带来的负面影响。Wu 等人[37]提出 Afdgcf(Adaptive feature decorrelation graph collaborative filtering)框架,设计自适应特征去相关图协同过滤框架,动态地将相关惩罚应用于嵌入矩阵的特征维度,有效缓解特征冗余问题。

可见,现有方法大多从缓解特征维度过度相关性的角度出发,提升模型的泛化能力。然而,这类方法普遍忽视其与过度平滑之间的潜在耦合关系,导致在深入建模用户偏好时难以全面优化特征表达,从而限制模型性能的进一步提升。

综上分析可知,一方面,现有基于注意力机制的 SR 方法往往忽视过度平滑问题在深层特征建模中的影响;另一方面,当前缓解特征过度相关性的方法未能充分考虑其与过度平滑之间的耦合效应,限制模型的泛化能力。基于此,为缓解过度平滑问题并提升模型泛化性,我们提出 MCMD-SR 模型。具体地,设计一种自注意力感知的特征掩码机制,构建注意力引导的掩码矩阵,抑制项目特征间的高相似性,增强模型对差异性特征的识别能力,同时引入对比学习框架,构造个性化表示样本对,提升模型对潜在个性偏好的建模能力,并与掩码机制协同缓解过度平滑现象;另外,设计一个多维度自适应去过度相关模块,从局部和全局两个层面建模特征间的相关性,动态过滤冗余信息,提高模型的泛化能力。

2 多策略融合推荐

问题定义:SR 是用户根据历史交互序列,预测下一时刻用户可能会选择的项目。假设所有的用户集 $U=\{u_1,u_2,\cdots,u_N\}$,项目集 $V=\{v_1,v_2,\cdots,v_M\}$。用户 u 的交互序列表示为 $S_u=<v_1,v_2,\cdots,v_t>$,用户 u 的交互时间步表示为 $T_u=\{1,2,\cdots,t\}$,给定用户 u 的历史交互序列 S_u,在候选项目集合 V 中寻找最可能作为用户下一步交互目标的项目 V_{t+1},如式(1)所示。

$$\arg\max_{v_k\in V} P(v_{t+1}=v_k|S_u) \tag{1}$$

其中:arg max 为使项目 v_k 概率取最大值的目标函数;$P(v_{t+1}=v_k|S_u)$ 表示在给定历史交互序列 S_u 的条件下,用户在下一时刻 $t+1$ 与项目 v_k 交互的条件概率。

为便于阐述,对文中一些主要符号进行说明,如表 1 所示。

表 1 主要符号及含义

Table 1 Main symbols and their meanings

符号	含义
$U=\{u_1,u_2\cdots,u_N\}$	用户集
$V=\{v_1,v_2,\cdots,v_M\}$	项目集
$S_u=<v_1,v_2,\cdots,v_t>$	用户 u 的交互序列
$T_u=\{1,2,\cdots,t\}$	用户 u 的交互时间步
$\boldsymbol{E}$	用户与项目的交互序列特征矩阵
l	特征矩阵的层数
$\boldsymbol{M}^l$	第 l 层的掩码矩阵
$g(\boldsymbol{M}^l)$	第 l 层掩码矩阵的掩码率
e_{mi}	特征矩阵 $\boldsymbol{E}$ 中第 m 个样本第 i 维度的嵌入
$\boldsymbol{E}_{*i}$	特征矩阵 $\boldsymbol{E}$ 的维度 i
$\bar{\boldsymbol{E}}_{*i}$	特征矩阵 $\boldsymbol{E}$ 维度 i 的特征平均值
e^l_{mi}	特征矩阵 $\boldsymbol{E}$ 第 l 层中样本 m 维度 i 的嵌入
$\overline{\boldsymbol{E}^l_{**}}$	特征矩阵 $\boldsymbol{E}$ 第 l 层的特征平均值
$\tilde{r}(\boldsymbol{E}^l)$	特征矩阵 $\boldsymbol{E}$ 第 l 层的列间平均相关性
$\gamma(\boldsymbol{E}^l)$	特征矩阵 $\boldsymbol{E}$ 第 l 层的列间相关性惩罚系数
$\phi(\boldsymbol{E}^l)$	特征矩阵 $\boldsymbol{E}$ 第 l 层的相关性
$\theta(\boldsymbol{E}^l)$	特征矩阵 $\boldsymbol{E}$ 第 l 层的相关性惩罚系数
λ	损失函数权重控制参数

一方面,目前缓解过度平滑问题主要在基于 GNN 的协同过滤推荐中受到关注,而在基于注意力机制的 SR 中却很少被提及;另一方面,基于注意力机制的 SR 模型未考虑降低项目特征维度相关性,使其泛化能力不强。基于此,提出一种 MCMD-SR 模型,通过构建自注意力特征掩码矩阵,以减少项目特征间的相似性;同时,利用对比学习使模型获得更多的个性化特征并结合自注意力掩码机制共同缓解过度平滑问题;另外,构建一个多维度自适应去过度相关模块,从局部和全局降低项目特征维度的过度相关性,减少特征冗余度,提升模型对关键特征的关注度和对项目的泛化性。图 1 是本文模型的总体框架,其工作原理如下:

(1)通过自注意力特征掩码减少项目间的相似特征,提高模型对差异项目特征的辨识性(见图1模块①)。利用自注意力掩码机制,有针对性地遮蔽部分特征,减少项目间的相似性,增强模型对差异性特征的关注,从而提高模型对差异项目的辨识性。

(2)利用对比学习,强化用户与项目的交互序列的个性特征(见图1模块①)。分别嵌入掩码矩阵的最后一层和初始特征矩阵的第一层,作为对比学习任务的正、负样本,通过最大化正样本相似性和最小化负样本相似性提升特征的区分性及嵌入表示的独特性。

(3)进行多维度自适应去过度相关,降低高维特征的相似性(见图1模块②)。对掩码后的高维特征嵌入矩阵分别计算列间和层间的皮尔逊相关(Pearson Correlation Coefficient,PCC)系数,根据系数自适应降低特征权重,减少特征冗余,提升模型对关键特征的关注度。

(4)通过多任务联合学习优化策略,旨在缓解过度平滑现象,减少特征冗余,并提升推荐结果的预测能力(见图1模块③)。具体而言,该框架将对比学习损失、自注意力掩码损失以及多维度自适应去相关损失进行联合优化。多任务联合优化不仅能够确保各损失项之间的协同作用,还能够增强特征表达的多样性与区分性,进一步提高推荐模型的泛化性能。

2.1　自注意力掩码机制与对比学习增强方法

在数据特征提取与建模过程中,针对有效挖掘项目的差异性并突出个性化特征这一问题,本节提出结合自注意力掩码机制与对比学习的增强策略,从局部特征的细化到全局特征的强化,极大地缓解过度平滑问题。自注意力掩码机制根据注意力分数对特征进行遮蔽,减少相似特征,从而使项目特征更具区分度;随着训练过程的深入,引入掩码权重衰减函数,充分发挥特征掩码在促进模型有效学习中的作用,从而增强差异性项目的辨识性。

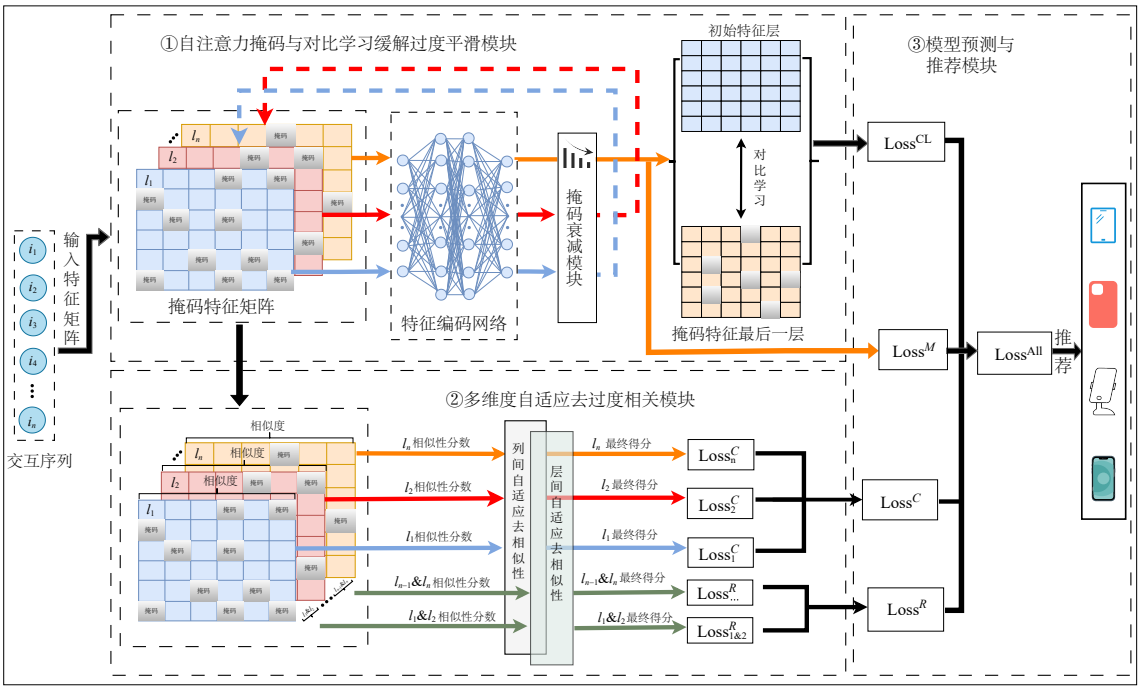


图1　序列推荐模型框架

Figure 1　The framework of the sequential recommendation model

对比学习则通过构建正负样本对以增强模型对个性特征的学习能力,利用最大化正样本对之间的相似性,同时最小化负样本对之间的相似性,使模型能捕捉到数据中潜在的个性化特征。自注意力掩码机制与对比学习相结合,模型既能有效捕捉项目差异性特征,又能增强对个性化特征的理解。

2.1.1　自注意力掩码机制增强项目特征辨识性

当项目特征经过多个网络层的学习后,基于自注意力机制的SR模型会融合历史时刻和当前时刻的特征,这往往使得项目特征之间变得难以区分,从而带来过度平滑的问题。

例1　用户u_1购买过手机钢化膜、手机保护壳、手机支架,基于用户交互序列,SR模型会抽取手机钢化膜特征,然后传递到下一时刻即抽取手机保护壳特征,同时与手机钢化膜的特征融合再传递到抽取手机支架特征。这些项目的共同特征是手机配件,模型根

据手机配件这一特征为用户推荐全为手机配件的项目。但因这些过度相似的特征导致推荐结果错误,故我们认为这些过度相似的特征应该屏蔽一部分,以增强项目的差异性。

由例1,本模块引入自注意力掩码机制,使项目特征呈现差异性,从而更有效地识别和捕捉高质量特征。bert(bidirectional encoder representations from transformers)[38]在输入文本序列中随机选择固定比例的单词,并将它们替换为特殊的掩码标记[MASK],然后基于上下文信息预测这些被掩码的单词。受bert启发,本文将掩码机制引入特征中,但目前掩码机制主要是对特征进行随机掩码并且采用固定的掩码比例,这既无法区分差异性特征,又无法动态调整掩码比例。基于此,提出自注意力掩码衰减机制,通过计算注意力分数对特征进行掩码,并根据网络层数的增加来相应地衰减掩码比例。利用自注意力掩码衰减机制,既突出了各特征的重要性又改变了掩码比例固定不变的不足,从而突出差异性项目。

我们通过自注意力机制计算特征矩阵中所有嵌入e_{mi}^{l}的相似度分数,如式(2)所示。

$$\text{score}(\boldsymbol{q}_i,\boldsymbol{k}_j)=\frac{\boldsymbol{q}_i\cdot\boldsymbol{k}_j}{\sqrt{d_k}} \tag{2}$$

其中:每个特征嵌入e_{mi}^{l}进行线性变换得到3个向量:查询向量$\boldsymbol{q}_i$、键向量$\boldsymbol{k}_i$和值向量$\boldsymbol{z}_j$;$\text{score}(\boldsymbol{q}_i,\boldsymbol{k}_j)$表示第$i$个查询向量$\boldsymbol{q}_i$与第$j$个键向量$\boldsymbol{k}_j$的相似度;$d_k$是键向量$\boldsymbol{k}$的维度,然后对这些得分进行softmax操作,得到归一化的权重α_{ij},如式(3)所示。

$$\alpha_{ij}=\frac{\exp\left(\text{score}(\boldsymbol{q}_i,\boldsymbol{k}_j)\right)}{\sum_{f=1}^{d}\text{score}\,(\boldsymbol{q}_i,\boldsymbol{k}_j)} \tag{3}$$

其中,d为特征嵌入的维度。最终,模型根据计算的注意力权重,对所有值$\boldsymbol{z}_j$进行加权求和,得到特征嵌入e_{mi}^{l}的注意力分数,如式(4)所示。

$$\text{sco}(e_{mi}^{l})=\sum_{j=1}^{n}\alpha_{ij}\boldsymbol{z}_j \tag{4}$$

其中,$\text{sco}(e_{mi}^{l})$反映了e_{mi}^{l}在该项目中的重要程度。基于特征重要性的不同,我们据此决定对哪些特征进行掩码。具体而言,计算特征矩阵中每个特征的样本均值μ_m,将其作为判定是否进行掩码的阈值。对低于该阈值的特征,采用掩码操作以减少无关特征的干扰,如式(5)所示。

$$\text{mask}_{ij}^{l}=\begin{cases}1, & \text{if sco}(e_{mi}^{l})<\mu_m\\ 0, & \text{otherwise}\end{cases} \tag{5}$$

其中:mask_{ij}^{l}表示掩码矩阵第l层中样本i在特征维度j上的掩码值;μ_m是第m个样本的掩码阈值。对于被掩码的特征,我们将其与上一层特征表示融合,构建新的嵌入表示,如式(6)所示。

$$\boldsymbol{E}^{l}=\boldsymbol{E}^{l}\cdot\boldsymbol{M}^{l}+\boldsymbol{E}^{l-1}\cdot(1-\boldsymbol{M}^{l}) \tag{6}$$

其中:$\boldsymbol{E}^{l}$为特征矩阵第l层;$\boldsymbol{E}^{l-1}$为特征矩阵第$l-1$层;$M^{l}\in\{0,1\}^{m\times d}$为掩码矩阵,其元素取值为0或1;符号“·”表示两个矩阵对应元素相乘。鉴于第一层特征表示通常蕴含丰富的信息且特征分布差异显著,这里选择从第二层特征矩阵开始应用特征掩码。为了在浅层施加更强的扰动以抑制早期的过度平滑,同时在深层仍保持适度的扰动以约束晚期的过度聚合,我们设计了随层数递减的对数掩码率函数,如式(7)所示。

$$g(\boldsymbol{M}^{l})=\begin{cases}0, & l<2\\ \log(\omega/l+1), & \text{otherwise}\end{cases} \tag{7}$$

其中:$g(\boldsymbol{M}^{l})$表示特征矩阵第l层的掩码率;$\omega>0$为控制衰减形状的超参数。与线性衰减(即$n(\boldsymbol{M}^{l})=a-bl$)或指数衰减(即$p(\boldsymbol{M}^{l})=a\cdot\mathrm{e}^{-bl}$)不同,这里的对数形式在浅层下降速度相对更快,而在中后层又会逐渐趋于平缓,该衰减策略具有以下优势:

(1)模型在靠近输入的浅层受到较强的掩码约束,有利于打破高度相似的局部特征,缓解序列表示在早期阶段的过度平滑;

(2)在更深层中,掩码率不会衰减得过快,仍保留适度的扰动,以持续约束深层特征的过度聚合,避免模型过于依赖少数局部显著特征;

(3)因$g(\boldsymbol{M}^{l})$在$\left(0,\log(\omega/l+1)\right)$区间内有界且随$l$单调递减,故训练过程中掩码率变化较为平滑,有助于保持优化过程的稳定性。

进一步将线性衰减、指数衰减以及对数衰减函数的模型在Beauty数据集(见3.1节)上进行实验对比,以验证式(7)的对数设计带来的优势,如表2所示。

表2 各衰减函数在Beauty数据集下的对比

Table 2 Comparison of different decay functions on Beauty

衰减函数	H@5	H@10	H@20	N@5	N@10	N@20
$n(\boldsymbol{M}^{l})$	0.070 6	0.095 8	0.129 7	0.050 5	0.058 6	0.067 2
$p(\boldsymbol{M}^{l})$	0.070 1	0.095 1	0.128 1	0.050 5	0.058 5	0.066 9
$g(\boldsymbol{M}^{l})$	**0.072 7**	**0.098 7**	**0.132 5**	**0.052 0**	**0.060 3**	**0.068 8**

注:HR@K和NDCG@K在表中分别简写为H@K、N@K;本文所选衰减函数的实验数据以加粗字体标记。用$n(\boldsymbol{M}^{2})=p(\boldsymbol{M}^{2})=g(\boldsymbol{M}^{2})$和$n(\boldsymbol{M}^{L})=p(\boldsymbol{M}^{L})=g(\boldsymbol{M}^{L})$求解线性衰减与指数衰减函数的$a$,$b$。

这里选用Beauty数据集,一方面考虑到其显著的稀疏性与较短的平均序列长度。在此类交互信息有限的数据分布下,随着网络层数的增加,模型极易产生过度平滑现象,这对正则化机制在深层的持续性提出了较高要求。另一方面能够直观地反映不同衰减

函数的差异:线性或指数衰减往往在深层导致掩码率过早趋于零,从而失去对过度平滑的抑制作用;而我们提出的对数衰减策略能够在深层网络中维持非零且适度的掩码扰动。因此,在 Beauty 数据集上的实验对比,能最显著地验证本方法在特征正则化方面的优势,有力支撑了对数函数“前强后稳”设计的合理性。

由表2可看出,在3种随层数递减的策略中,对数衰减在各项推荐指标中均取得了最优值,说明“浅层较强、深层较弱”的掩码约束最有效。

2.1.2　对比学习强化项目个性特征

通过自注意力掩码衰减机制,可有针对性地屏蔽部分特征,从而引导模型学习更多差异化的特征表示,增强项目之间特征的差异性。然而,个性化特征仍然不够显著。为此,我们期望借助对比学习进一步强化个性化特征,从而更加突出项目的差异化表示。

对比学习涉及计算正负样本对之间的相似度,若逐层执行此类计算,将显著增加模型的计算开销。为优化效率,我们仅在特征矩阵的初始层和经过自注意力掩码机制处理的最终层上计算对比损失,该策略可显著降低计算复杂度。此外,模型的最后一层特征通常已综合了底层和中间层的语义信息。直接在该层进行对比学习,可有效提升模型对高层语义的区分能力,从而进一步突显不同项目的个性化特征。对比学习通过利用正负样本对,来优化特征表示的学习。通过最小化对比损失,模型能够将正样本对的特征距离拉近,同时将负样本对的特征距离推远,从而提升特征的区分能力。

例2　用户 u_1 购买过手机钢化膜、手机保护壳、手机支架等配件,这些配件通常是为了购买手机而准备的。在嵌入空间中,手机配件的嵌入表示应当被适当拉远,而手机本身的嵌入表示则应被拉近,以更准确地反映用户的主要兴趣和个性化偏好。

选择原始特征矩阵的第一层以及经过自注意力掩码机制处理的特征矩阵最后一层作为对比学习的两个样本。低层特征通常包含更为基础的个性化信息,如项目嵌入的基本属性;而经过自注意力掩码机制后的高层特征能够基于这些基础特征进行上下文建模,捕捉到更加复杂和细粒度的个性化模式,如项目在特定场景下的表现差异。这种对比方式有助于模型更好地学习到差异化的特征表示,从而进一步突显个性化特征,如式(8)所示。

$$\mathrm{Loss}^{\mathrm{CL}}=-\sum_{m=1}^{M}\sum_{i=1}^{d}\log\left(\frac{\exp\left(\frac{\mathrm{sim}(e_{mi}^{1},e_{mi}^{L})}{\tau}\right)}{\sum\limits_{\substack{n=1\\(n,j)\neq(m,i)}}^{M}\sum\limits_{j=1}^{d}\left[\exp\left(\frac{\mathrm{sim}(e_{mi}^{L},e_{nj}^{L})}{\tau}\right)+\exp\left(\frac{\mathrm{sim}(e_{mi}^{1},e_{nj}^{L})}{\tau}\right)\right]}\right)\tag{8}$$

其中:m 和 n 分别为样本索引;M 是样本总数;i 和 j 表示特征索引;L 表示特征层总数;$\mathrm{sim}(\cdot)$ 计算两个特征之间的相似性;e_{mi}^{1} 表示矩阵中原始特征层的嵌入表示;e_{mi}^{L} 则表示经过自注意力掩码机制处理的特征矩阵最后一层中的嵌入表示。$\mathrm{sim}(e_{mi}^{1},e_{mi}^{L})$ 用于计算同一特征在不同层之间的相似性,构成正样本对;而 $\mathrm{sim}(e_{mi}^{L},e_{nj}^{L})$ 以及 $\mathrm{sim}(e_{mi}^{1},e_{nj}^{L})$ 则用于计算同层或跨层的不同特征之间的相似性,构成负样本对。$\exp(\cdot)$ 将相似度值转化为概率分布,以增强对比学习中的梯度可传递性。温度参数 τ 用于调节相似性分布的平滑度,以控制相似度的敏感性。

2.2　多维度自适应去过度相关方法

在高维特征空间中,特征间的高度相关性会降低模型的泛化性能,尤其在处理复杂交互数据时更为显著。因此,缓解高维特征间的过度相关性是提升模型泛化性的关键环节。为此,我们引入一种多维度自适应的去相关方法,旨在降低特征之间的过度相关性,减少冗余特征信息,进而提升模型的泛化能力和在推荐任务中的表现。

在应用自注意力掩码机制对特征进行处理后,计算特征向量间的PCC系数。基于计算所得的PCC值,引入一项惩罚项,对于高PCC值的特征对,施加较大的惩罚强度;反之,对于低PCC值的特征对,则施加较小的惩罚强度。该动态惩罚机制旨在降低高维特征间的冗余性,从而提升对特征的判别能力。尽管列间去过度相关能够在一定程度上降低高维特征的相关性,但仅从列间角度进行去相关处理并不足以全面降低高维特征的整体相关性。这是因为列间去过度相关主要关注于特征之间的依赖关系,而忽略了多层结构中不同层之间由于信息交互而产生的复杂依赖关系。

进一步采用层间去相关方法,以协同抑制高维特征空间中冗余相关性。在特征维度分析中,通过计算特征维度层级间PCC系数,构建动态惩罚调节机制。即,对相关性系数较高的特征维度层施以更高惩罚权重,对相关性较低的特征层则相应降低惩罚力度,通过自适应调节层间特征的惩罚系数实现维度间相关性的有效抑制。

这样，特征列间去过度相关与特征层级间去过度相关形成互补的双重去相关策略，前者聚焦于同一特征空间内列维度的局部相关性消解，后者致力于跨层次特征维度的全局相关性优化。二者协同作用可实现高维特征空间中冗余信息的多层次过滤，显著提升特征表示的稀疏性与判别力，进而增强模型在跨场景推荐任务中的泛化能力与推荐性能。

常见的相关性度量方法包括距离相关性（Distance Correlation，DCor）、余弦相似度（COsine Similarity，COS）、欧几里得距离（Euclidean Distance，ED）以及皮尔逊相关系数（PCC）。尽管DCor、COS、ED、PCC等方法均可用于衡量特征维度间的相关性，但它们在本任务中的适配性、计算效率与稳定性存在明显差异。

首先，从适配性角度来看，鉴于本任务中需对特征矩阵任意两列间和相邻层进行大量相关性计算，而SR模型中的特征在经过层归一化、嵌入表示等步骤后，分布已被标准化，使得不同特征维度之间主要呈现线性或准线性相关结构，而此时PCC能够更有效地衡量特征维度间的线性依赖关系。相比之下，DCor过度关注非线性依赖而带来冗余计算，故与本任务适配性较弱；COS虽然在嵌入表示中应用广泛，但其度量的是向量方向相似性，而非统计意义上的相关性，在本任务中适应性也较弱；ED主要用于度量样本之间的距离，在高维特征空间中，使得距离度量不再具有良好的区分能力，无法有效识别特征冗余，同样在本任务中适应性较弱。

其次，从计算效率的角度来看，PCC的复杂度为$O(d^2M)$，仅依赖特征矩阵的均值与方差即可计算，适合在多层网络结构中频繁调用；而距离相关性DCor需计算样本间距离矩阵，其复杂度通常达到$O(M^2)$，在M较大时会显著增加模型训练时间，使该方法难以在深层序列推荐模型中得到高频应用。COS与ED在计算特征维度间的相关性时，其总体时间复杂度均为$O(d^2M)$，虽与PCC处于同一量级，但COS度量向量的方向相似性、ED度量几何距离，它们均不具有统计意义上的相关性。综合来看，在高维特征的维度相关性度量中，DCor、COS、ED、PCC等方法的计算代价存在明显差异，然而PCC兼具较低的计算成本与较好的统计相关性，故在本任务中使用更为适合。

最后，从稳定性的角度来看，PCC作为一种相对轻量的度量方法，在面对标准化后的特征时，表现出较强的稳定性。相比之下，DCor、COS和ED在计算过程中受噪声的影响较大，导致它们不如PCC稳定。因此，在要求计算结果稳定的本任务中，PCC更为可靠。

为进一步阐明列间、层间自适应去过度相关均用PCC的优势，我们在LastFM数据集（见3.1节）上通过实验对比了DCor、COS、ED、PCC的任意组合表现，如表3所示。

表3 LastFM数据集上列间与层间自适应去过度相关的不同组合相关性度量方法结果对比

Table 3 Comparison of the results of different combinations of correlation metrics for column-wise and layer-wise adaptive decorrelation on LastFM

相关性度量方法		H@5	H@10	H@20	N@5	N@10	N@20
列间自适应去过度相关	层间自适应去过度相关						
DCor	COS	0.052 3	0.068 8	0.098 2	0.036 3	0.041 5	0.048 9
DCor	ED	0.048 6	0.073 4	0.101 8	0.035 8	0.043 7	0.050 8
DCor	PCC	0.056 9	0.071 6	0.103 7	0.037 5	0.042 1	0.050 1
DCor	DCor	0.047 7	0.063 3	0.101 8	0.033 7	0.038 8	0.048 3
COS	DCor	0.049 5	0.059 6	0.095 4	0.035 5	0.038 8	0.047 7
COS	ED	0.045 9	0.059 6	0.098 2	0.032 8	0.037 4	0.046 9
COS	PCC	0.052 3	0.067 9	0.099 1	0.037 6	0.042 5	0.050 3
COS	COS	0.046 8	0.063 3	0.103 7	0.032 2	0.037 5	0.047 5
ED	DCor	0.048 6	0.071 6	0.101 8	0.033 7	0.041 0	0.048 6
ED	COS	0.045 0	0.067 9	0.101 8	0.034 2	0.041 4	0.050 0
ED	PCC	0.052 3	0.071 6	0.105 5	0.035 8	0.042 0	0.050 5
ED	ED	0.051 4	0.072 5	0.110 1	0.036 5	0.043 4	0.052 8
PCC	DCor	0.049 5	0.070 6	0.106 4	0.035 6	0.042 3	0.051 3
PCC	COS	0.051 4	0.072 5	0.099 1	0.035 7	0.042 4	0.049 0
PCC	ED	0.050 5	0.067 9	0.106 4	0.034 7	0.040 3	0.049 9
PCC	**PCC**	**0.051 4**	**0.075 2**	**0.111 9**	**0.037 4**	**0.045 1**	**0.054 2**

注：HR@K和NDCG@K在表中分别简写为H@K、N@K；本文所选相关性度量方法组合的实验数据以加粗字体标记，为便于比较，利用下划线突显对比相关性度量方法组合中的最佳数据。

这里选用LastFM数据集,首先,其用户行为序列具有典型的中等长度特征(平均长度48.2),既包含连续的短期偏好,又保留足够的历史信息,使特征维度间更易累积冗余与产生趋同;其次,其序列波动性强、音乐项目语义相似度高,特征嵌入在多层传播后更易增强相关性;最后,中等序列长度避免了超长序列的噪声累积与短序列的特征不足,使不同相关性度量的比较更加客观稳定。因此,在LastFM数据集上进行对比实验,能有效验证多维度去过度相关均使用PCC方法在去特征冗余方面的优势。由表3可看出,多维度去过度相关均使用PCC方法在H@10、H@20、NDCG@10、NDCG@20指标中均取得最优值。虽然在H@5和NDCG@5指标上未获得最优值,但是与最优值相差仅为0.005 5和0.000 2,差距较小。整体来看,多维度去过度相关均使用PCC方法推荐效果最佳,将其作为相关性度量方式最有效。

综上分析,多维度去过度相关均采用PCC方法作为相关性度量方式,是兼顾适配性、计算高效性、模块协同稳定性及有效性的最佳选择。

2.2.1　列间自适应去过度相关

在序列推荐的特征嵌入矩阵中,不同的列向量通常代表项目在潜在空间中的不同属性维度。然而,随着深度神经网络的训练,特征维度间往往会出现高度的共线性,即不同的维度编码了重复或冗余的语义信息。这种维度的过度相关不仅浪费有限的表达空间,还易导致模型对特定特征的过拟合。为获取特征维度间的线性依赖关系并据此进行去冗余处理,采用PCC来衡量任意两列特征间的相关程度。基于PCC构建的自适应惩罚机制,能识别并抑制高相关性的特征维度,促使模型学习到更加独立且紧凑的特征表示。

例3　用户u_1购买过手机钢化膜、手机保护壳和手机支架,这些项目的共同特征之一是“手机配件”,另一共同特征是“手机周边”。“手机配件”和“手机周边”之间存在较高的相关性,会导致特征冗余。

为此,在用PCC来计算相关性系数时,对高度相关的特征施加更大的惩罚权重,而对于低相关性的特征则赋予较小的惩罚权重,从而实现特征的自适应优化,如式(9)所示。

$$r^l(\boldsymbol{E}^l_{*i},\boldsymbol{E}^l_{*j})=\frac{\mathrm{cov}(\boldsymbol{E}^l_{*i},\boldsymbol{E}^l_{*j})}{\sigma_{\boldsymbol{E}^l_{*i}}\sigma_{\boldsymbol{E}^l_{*j}}} \tag{9}$$

其中:$\boldsymbol{E}^l_{*i}$和$\boldsymbol{E}^l_{*j}$分别表示特征矩阵$\boldsymbol{E}$第l层的第i,第j个特征维度;cov(·)用于计算两特征维度间的协方差,如式(10)所示。

$$\mathrm{cov}(\boldsymbol{E}^l_{*i},\boldsymbol{E}^l_{*j})=\frac{\sum_{m=1}^{M}\left(e^l_{mi}-\overline{e^l_{*i}}\right)\left(e^l_{mj}-\overline{e^l_{*j}}\right)}{M} \tag{10}$$

其中:e^l_{mi}和e^l_{mj}分别表示特征矩阵在l层中维度i和j上,样本m的值;$\overline{e^l_{*i}}$和$\overline{e^l_{*j}}$则分别是对应的特征均值。协方差用于描述特征维度i和j的共同变化程度,反映了它们的相关性。$\sigma_{\boldsymbol{E}^l_{*i}}$和$\sigma_{\boldsymbol{E}^l_{*j}}$分别为对应维度特征的标准差,如式(11)所示。

$$\begin{cases}\sigma_{\boldsymbol{E}^l_{*i}}=\sqrt{\dfrac{\sum_{m=1}^{M}\left(e^l_{mi}-\overline{e^l_{*i}}\right)^2}{M}}\\ \sigma_{\boldsymbol{E}^l_{*j}}=\sqrt{\dfrac{\sum_{m=1}^{M}\left(e^l_{mj}-\overline{e^l_{*j}}\right)^2}{M}}\end{cases} \tag{11}$$

这表示在特征矩阵第l层中,第i列特征维度$\boldsymbol{E}^l_{*i}$和第j列特征维度$\boldsymbol{E}^l_{*j}$的离散波动性。通过列间PCC系数运算,我们可得到第l层中特征维度i和维度j的PCC相关系数$r^l(\boldsymbol{E}^l_{*i},\boldsymbol{E}^l_{*j})$。基于该度量,本文将降低嵌入矩阵列间的平均相关性作为核心优化目标,如式(12)所示。

$$\bar{r}(\boldsymbol{E}^l)=\frac{\sum_{i\neq j}r^l(\boldsymbol{E}^l_{*i},\boldsymbol{E}^l_{*j})^2}{d(d-1)},\quad i,j\in\{1,2,\cdots,d\} \tag{12}$$

其中,$\bar{r}(\boldsymbol{E}^l)$是特征矩阵在$l$层中各特征维度之间的PCC系数平方的均值。若特征维度之间存在较高相关性,则说明它们存在一定冗余。平方加权的方式能够放大高相关特征的影响,促使模型学习到更加解耦、差异化的特征表示。将$\bar{r}(\boldsymbol{E}^l)$作为衡量第$l$层特征矩阵中列间相关性的指标,旨在动态调整对不同相关性特征的惩罚权重。在保持总惩罚强度不变的前提下,我们对高度相关的特征维度分配更大的惩罚权重,而对低相关性的特征维度施加较小的惩罚,如式(13)所示。

$$\gamma(\boldsymbol{E}^l)=\frac{1/\bar{r}(\boldsymbol{E}^l)}{\sum_{l=1}^{L}1/\bar{r}(\boldsymbol{E}^l)} \tag{13}$$

其中,l为当前特征层;L为总特征层数。

因此,可获得列间自适应去过度相关损失Loss^c,如式(14)所示。

$$\mathrm{Loss}^c=\sum_{l=1}^{L}\bar{r}(\boldsymbol{E}^l)\cdot\gamma(\boldsymbol{E}^l) \tag{14}$$

2.2.2　层间自适应去过度相关

通过列间自适应去过度相关,可从列间的视角对特征相关性进行惩罚,缓解一部分高维特征的过度相关,但列间自适应去过度相关忽略了不同层次之间的

相关性，不能充分减少高维特征中的相关性和冗余，故结合层间自适应去过度相关方法来共同缓解高维特征中的过度相关性。

例 4 用户 u_1 购买过手机钢化膜、手机保护壳、手机支架，模型在学习项目特征时，底层特征较泛化，高层特征更抽象，这些项目的底层共同特征之一是小物件，而更高层特征之一是手机配件，小物件与手机配件过度相关。

对经过自注意掩码后的特征矩阵的相邻层计算相关性，将相关性高的特征层分配更大的惩罚系数，将相关性小的特征层赋予更小惩罚权重进行学习优化。与列间不同，层间自适应去过度相关从特征层整体角度考虑它们之间相关性。特征层 l 的 PCC 系数 $\phi(\boldsymbol{E}^l)$ 计算，如式(15)所示。

$$\phi(\boldsymbol{E}^l)=\begin{cases}\dfrac{\operatorname{cov}(e_{mi}^1,e_{mi}^2)}{\sigma_{e_{mi}}^1\cdot\sigma_{e_{mi}}^2}, & \text{if } l=1\\ \dfrac{\operatorname{cov}(e_{mi}^{l-1},e_{mi}^l)}{\sigma_{e_{mi}}^{l-1}\cdot\sigma_{e_{mi}}^l}\cdot 0.5+\dfrac{\operatorname{cov}(e_{mi}^l,e_{mi}^{l+1})}{\sigma_{e_{mi}}^l\cdot\sigma_{e_{mi}}^{l+1}}\cdot 0.5, & \text{if } 1<l<L\\ \dfrac{\operatorname{cov}(e_{mi}^{L-1},e_{mi}^L)}{\sigma_{e_{mi}}^{L-1}\cdot\sigma_{e_{mi}}^L}, & \text{if } l=L\end{cases} \tag{15}$$

其中：e_{mi}^l，e_{mi}^{l+1} 分别为特征矩阵 l 和 $l+1$ 层在样本 m 维度 i 上的值；$\operatorname{cov}(\cdot)$ 为特征矩阵 l 和 $l+1$ 层的协方差，如式(16)所示。

$$\operatorname{cov}(e_{mi}^l,e_{mi}^{l+1})=\frac{\sum_{m=1}^{M}\sum_{i=1}^{d}\left(e_{mi}^l-\overline{e_{**}^l}\right)\left(e_{mi}^{l+1}-\overline{e_{**}^{l+1}}\right)}{M} \tag{16}$$

其中，$\overline{e_{**}^l}$ 和 $\overline{e_{**}^{l+1}}$ 分别为特征矩阵 l 和 $l+1$ 层的特征平均值。通过协方差表示特征矩阵 l 和 $l+1$ 层特征的共同变化程度 $\sigma_{e_{mi}^l}$、$\sigma_{e_{mi}^{l+1}}$ 为特征矩阵 l 和 $l+1$ 的标准差，如式(17)所示。

$$\begin{cases}\sigma_{e_{mi}^l}=\sqrt{\dfrac{\sum_{m=1}^{M}\sum_{i=1}^{d}\left(e_{mi}^l-\overline{e_{**}^l}\right)^2}{M}}\\ \sigma_{e_{mi}^{l+1}}=\sqrt{\dfrac{\sum_{m=1}^{n}\sum_{i=1}^{d}\left(e_{mi}^{l+1}-\overline{e_{**}^{l+1}}\right)^2}{n}}\end{cases} \tag{17}$$

这表示特征矩阵 l 和 $l+1$ 层的离散程度，反映每层特征的波动性。经过计算特征层 PCC 系数，可得到特征矩阵 l 和 $l+1$ 层的 PCC 相关系数 $\phi(\boldsymbol{E}^l)$。我们根据特征矩阵层 PCC 系数动态地对相应的特征层进行相关性惩罚，如式(18)所示。

$$\theta(\boldsymbol{E}^l)=\begin{cases}\dfrac{1/\phi(\boldsymbol{E}^1)}{\sum_{h=1}^{L}\left(1/\phi(\boldsymbol{E}^h)\right)}, & \text{if } l=1\\ \dfrac{1/\left(\phi(\boldsymbol{E}^{l-1})\cdot 0.5+\phi(\boldsymbol{E}^l)\cdot 0.5\right)}{\sum_{h=1}^{L}\left(1/\phi(\boldsymbol{E}^h)\right)}, & \text{if } 1<l<L\\ \dfrac{1/\phi(\boldsymbol{E}^L)}{\sum_{h=1}^{L}\left(1/\phi(\boldsymbol{E}^h)\right)}, & \text{if } l=L\end{cases} \tag{18}$$

对各特征层进行不同的惩罚系数计算。注意，因计算的是相邻层相关系数，故对 $1<l<L$ 的重叠特征层，相关性惩罚系数需综合相邻两层进行计算。若不存在明显的特征偏向，则将权重均分能有效平衡邻接层的影响，避免单层特征在融合过程中主导模型，导致信息冗余或丢失，故将相邻层的相关性权重各取 0.5。层间自适应特征去相关损失 Loss^R，如式(19)所示。

$$\text{Loss}^R=\sum_{l=1}^{L}\phi(\boldsymbol{E}^l)\cdot\theta(\boldsymbol{E}^l) \tag{19}$$

2.3 模型预测与推荐

本节给出模型 MCMD-SR 的整体预测过程，并详细阐述其整体推荐流程 。

2.3.1 模型预测

在第 2.1 节，我们基于自注意力掩码机制计算注意力分数，并据此对原始特征矩阵进行掩码操作，获得掩码处理后的特征矩阵。因 BSARec(Beyond Self-Attention for sequential Recommendation)[1]模型反映了序列推荐中普遍存在的过度平滑问题，是一个具有代表性的最新架构，故将该掩码矩阵输入至基础模型 BSARec 进行特征编码，最终计算掩码损失 Loss^M。

考虑到个性化特征的不显著性可能加剧过度平滑问题，我们进一步增强掩码处理后的最后一层特征，借助对比学习强化个性化特征，得到损失 Loss^{CL}。

尽管从特征维度的行方向上缓解过度平滑问题取得成效，但在提升模型的泛化能力和准确性方面，特征的高维过度相关性仍然不可忽视。高维特征的过度相关性易导致模型学习到冗余信息，进而削弱其泛化能力。故在第 2.2 节，我们从数据特征间的直接依赖关系出发，引入一种自适应列间去过度相关机制，得到损失 Loss^C。此外，为更全面地降低高维特征中的冗余性，我们进一步从模型不同层次之间的特征相关性出发，进行自适应层级间去过度相关，得到损失 Loss^R。据此得到模型的总损失 Loss^{All}，如式(20)所示。

$$\text{Loss}^{\text{All}}=\text{Loss}^{M}+\lambda_1\text{Loss}^{\text{CL}}+\lambda_2\text{Loss}^{C}+\lambda_3\text{Loss}^{R} \quad (20)$$

其中:λ_1 为对比学习损失权重;λ_2 与 λ_3 分别为控制列间、层间自适应去过度相关权重超参数。

2.3.2　推荐算法

我们通过构建自注意力掩码特征矩阵并对其进行对比学习,以缓解过度平滑问题,且联合列间、层间自适应去过度相关策略,达到提升模型泛化能力的目标。具体过程见算法1。

算法1　特征掩码与对比学习融合多维度去过度相关算法

输入:用户交互序列 S_u;
输出:项目列表List;
Begin
1. 初始化用户交互序列的嵌入矩阵 $\boldsymbol{E}$;
/* 自注意力掩码机制与对比学习 */
2. **For** l **in** $\boldsymbol{E}$
3. $\text{sco}(e^l_{mi})$←e^l_{mi} 嵌入的注意力分数;//见式(2)~(4)
4. **If** $l\geqslant 2$ **then**
5. 计算特征矩阵中各维度特征均值 μ_i;
6. **If** $\text{sco}(e^l_{mi})<\mu_i$ **then**
7. $\boldsymbol{E}^l$←特征掩码后的特征层; //见式(5)~(6)
8. **End If**
9. 衰减掩码率; //见式(7)
10. **End If**
11. **If** $l=L$ **then**
12. Loss^{CL}←$\boldsymbol{E}^1$ 层与 $\boldsymbol{E}^L$ 层对比学习的损失;//见式(8)
13. **End If**
14. **End For**
15. Loss^{M}←特征矩阵 $\boldsymbol{E}^1$ 的损失; //见文献[18]
/* 多维度自适应去过度相关 */
16. **For** l **in** $\boldsymbol{E}$
17. $r^l(\boldsymbol{E}^l_{*i},\boldsymbol{E}^l_{*j})$←特征矩阵第 l 层中维度 $\boldsymbol{E}^l_{*i}$ 和 $\boldsymbol{E}^l_{*j}$ 的PCC系数;
//见式(9)~(11)
18. $\bar{r}(\boldsymbol{E}^l)$←特征矩阵第 l 层列间相关性的平均值; //见式(12)
19. $\gamma(\boldsymbol{E}^l)$←特征矩阵第 l 层列间相关性的惩罚系数; //见式(13)
20. Loss^{c}←列间去过度相关总损失; //见式(14)
21. $\phi(\boldsymbol{E}^l)$←特征矩阵第 l 层的PCC系数;//见式(15)~(17)
22. $\theta(\boldsymbol{E}^l)$←特征矩阵第 l 层的层间相关性惩罚系数; //见式(18)
23. Loss^{R}←层间去过度相关总损失;//见式(19)
24. **End For**
25. 计算总损失 Loss^{All} 并更新参数; //见式(20)
26. **Output** 项目列表List
End

在算法1中,第1行初始化用户交互序列的嵌入矩阵 $\boldsymbol{E}$;第2~10行计算掩码并得到掩码后的特征层;第11~14行对掩码矩阵进行对比学习并计算对比学习损失;第15行计算特征掩码矩阵损失;第16~20行计算列间自适应去过度相关及列间总损失;第21~24行计算层间自适应去过度相关及层间总损失;第25行计算总损失 Loss^{All} 并更新参数;第26行输出推荐列表。

3　实验及其分析

为验证MCMD-SR模型的有效性,本文选择4个经典数据集Beauty、Yelp、LastFM、ML-1M进行综合实验对比分析,重点回答以下几个问题。

(1)RQ1:与经典的、最新的模型相比,本文模型的推荐效果如何?

针对此问题,3.4.1节设置对比实验,分别将本文模型MCMD-SR与11种相关模型作对比。实验结果表明,在HR@5、HR@10、HR@20、NDCG@5、NDCG@10、NDCG@20这6个评价指标上,模型MCMD-SR均优于其他对比模型。可知,所提模型在缓解过度平滑和降低特征相关性方面优势明显,能提升推荐性能和泛化性。

(2)RQ2:模型各构件是否有存在的必要性?

针对此问题,3.4.2节设置了消融实验,根据自注意力掩码机制、对比学习模块、列间自适应去过度相关、层间自适应去过度相关这4个构件组成4种变体模型,分析这些构件对模型推荐性能的影响。实验结果表明,这4种变体模型的推荐效果均在不同程度上劣于所提模型MCMD-SR,说明这4个构件在缓解过度平滑问题和降低特征相关性上具有一定效果。

(3)RQ3:主要超参数对模型效果有何影响?

针对此问题,3.4.3节设置了参数敏感度实验,重点选择对比学习损失权重、列间与层间自适应去过度相关权重展开分析,发现它们在不同的数据集上使模型达到最优时的取值不同,并分析了其潜在的原因,以便更好地优化模型。

所提模型MCMD-SR基于PyTorch框架,在Windows1064位操作系统,PyCharm 2019编辑器,和Python 3.9环境下实现。实验配置为32 GB内存,AMD R7 3700X 3.6 GHz CPU以及Nvidia GeForce RTX 2080Ti GPU。模型的嵌入向量维度为64,训练批次大小为256,学习率在{0.000 5,0.001,0.005}中选取,epoch为200,网络层数 l 在{1,2,4,6,8}中查找,对比学习损失权重 λ_1 在{0.000 1,0.000 5,0.01,0.05,1}上选取,列间和层间自适应去过度相关的损失权重 λ_2,λ_3 分别在{0.000 1,0.000 2,0.01,0.02,1}和{0.000 01,0.000 02,0.001,0.1,0.2}中搜索。训练中采用早停策略防止过拟合,优化器为Adam,最大序列长度设为50。

3.1　数据集

本文选取4个来自真实业务场景且近年均被大量学者使用和研究的经典数据集,分别是Beauty

(https://jmcauley. ucsd. edu/data/amazon/)、Yelp(https://www.yelp.com/dataset)、LastFM(https://grouplens.org/datasets/hetrec-2011/)、ML-1M(https://grouplens.org/Datasets/ movIelens/)。它们涵盖美容、美食、音乐和电影等多个重要领域,能有效反映多样化的用户需求和偏好,可全面验证推荐模型在不同用户行为模式及序列特性下的性能表现。统计结果如表4所示。

表4 数据集信息

Table 4 Dataset information

数据集	用户数	项目数	交互数	平均长度	稀疏性
Beauty	22 363	12 101	198 502	8.9	99.93%
Yelp	30 431	20 033	316 354	10.4	99.95%
LastFM	1 090	3 646	52 551	48.2	98.68%
ML-1M	6 041	3 417	999 611	165.5	95.16%

(1)Beauty:为美妆产品评论数据集,来自Amazon平台,涵盖护肤、化妆、香水、美发等多个美妆类产品,包含用户的文本评论、评分以及时间戳等信息,能真实反映用户对美妆产品的使用反馈。

(2)Yelp:由Yelp官方公开发布的餐饮评价数据集,包含商户信息、用户评论、评分以及用户交互行为(如签到、收藏)等数据。

(3)LastFM:为推荐系统领域经典的音乐流媒体用户行为数据集,广泛应用于序列推荐研究,采集自Last.fm平台的真实用户交互日志,完整记录了用户与音乐项目(歌曲/艺术家)的时序交互序列。

(4)ML-1M:为推荐系统研究领域最常用的数据集之一,包含用户基本信息、电影类型、上映时间及用户对电影的星级评价等多维度信息。

3.2 评价指标

在序列推荐中,命中率(Hit Ratio,HR)和归一化折损累积增益(Normalized Discounted Cumulative Gain,NDCG)是评估Top-K推荐效果的两个核心指标,共同衡量推荐模型生成的前K个项目是否准确反映用户的兴趣偏好。

(1)HR:主要用于衡量推荐列表中是否命中用户真正喜欢的项目,即推荐的覆盖率,如式(21)所示。

$$\mathrm{HR}=\frac{1}{|U|}\sum_{i=1}^{|U|}h_i \tag{21}$$

其中:$|U|$为用户数;h_i表示第i个用户访问的项目是否在推荐列表中,是则取值为1,表示“击中”,否则为0,表示“未击中”。

(2)NDCG:它关注的是这些项目在推荐列表中的排名,主要衡量推荐项目的排序质量,优先推荐排名较高的项目,以确保用户能够更快地接触到最相关的内容,如式(22)所示。

$$\mathrm{NDCG}=\sum_{u\in U}\frac{1}{Y_u}\sum_{i=1}^{K}\frac{\mathrm{rel}_i}{\log_2(i+1)} \tag{22}$$

其中:Y_u为用户u的真实值;k为已推荐项目数;rel_i是排名第i个物品的相关性分数。

HR主要用于衡量推荐模型对用户感兴趣内容的覆盖程度,而NDCG则侧重于评估推荐结果的排序是否符合用户的实际偏好。这两个指标相辅相成,可从覆盖率与排序质量两方面全面反映推荐模型的性能。指标值越高,说明推荐模型在该数据集上表现越佳。

3.3 对比模型

本节将所提模型MCMD-SR与当前较流行的几类序列推荐模型进行对比实验及分析,包括基于卷积神经网络(Convolutional Neural Network,CNN)或循环神经网络(Recurrent Neural Network,RNN)、基于自注意力机制、基于对比学习及特征增强、基于频域建模缓解过度平滑,以及基于增强视图优化对比学习的SR模型,以阐明本文模型的优势。

对于这5种对比模型,①在基于CNN或RNN的SR模型中,Caser(Convolutional sequence embedding recommendation model)[39]相比于传统CNN能更高效地提取特征,同时避免梯度消失问题。GRU4Rec[21]则利用RNN处理用户交互数据,并在一定程度上减轻RNN的计算负担。②在基于自注意力机制的SR模型中,SASRec[22]不同于传统基于RNN或CNN的方法,完全依赖自注意力机制进行序列建模,未使用RNN/CNN。BERT4Rec(sequential recommendation with Bidirectional Encoder Representations from Transformer)[40]则在多头自注意力机制的基础上引入掩码机制进行训练,以更好地学习用户的序列交互模式。③在基于对比学习及特征增强的SR模型中,DuoRec[30]结合对比学习和Transformer架构,从提高对噪声数据的鲁棒性角度提升推荐效果。FEARec(Frequency Enhanced hybrid Attention network for sequential Recommendation)[4]从用户行为特征增强的角度,提升用户偏好的表示能力。④在基于频域建模缓解过度平滑的SR模型中,FMLP-Rec(Filter-enhanced MLP is all you need for sequential Recommendation)[3]采用隐式策略去除数据噪声,在一定程度上缓解过度平滑问题。BSARec[1]使用傅里叶变换来调整高频和低频信息,让模型的注意力关注到更多高频信息而降低对低频的关注度,从而缓解注意力机制中的过度平滑问题。⑤在基于增强视图优化对比学习的SR模型中,CaDiRec[5]、MSDCCL[6]、FENRec[7]是最新的增强视图优化结合对比学习的SR模型。

(1)基于CNN或RNN的SR模型

该类模型通过建模用户行为序列的局部时序模

式或顺序依赖性,实现对用户短期兴趣的捕捉与预测。

①Caser:一种基于CNN的SR模型,利用CNN提取用户历史行为中的时序模式,预测下一个交互项。

②GRU4Rec:一种基于门控循环单元(Gated Recurrent Unit,GRU)的推荐模型方法,通过RNN捕捉用户的短期兴趣,通过门控机制可减轻该模型计算负担,提高训练效率和性能。

(2)基于自注意力机制的SR模型

这类模型旨在克服传统CNN或RNN在处理长距离序列依赖关系上的局限性,通过其固有的全局依赖建模能力,自注意力机制能更有效地捕捉用户长期偏好演化模式。

③SASRec:基于Transformer架构,利用自注意力机制有效捕捉用户的序列行为,特别是在处理远距离依赖关系时表现出更强建模能力。

④BERT4Rec:首次将BERT框架引入到SR任务中,采用双向Transformer结构,并使用掩码机制进行训练,能综合利用序列的前后文信息,从而更好地学习用户的序列交互模式。

(3)基于对比学习及特征增强的SR模型

这类模型通过多视角或多特征的信息对比与增强,旨在提升模型的表示能力和鲁棒性,可更好地适应噪声和多样化的用户行为,从而提高推荐效果。

⑤DuoRec:一种将对比学习与Transformer融合的SR模型,通过构建不同的样本视图来学习健壮的用户和项目表示,以增强模型对数据噪声的鲁棒性。

⑥FEARec:一种基于特征增强和注意力机制的模型,通过融合用户的多种行为序列特征来学习更丰富的表征,进而提升推荐性能。

(4)基于频域建模缓解过度平滑的SR模型

这类模型旨在利用频域分析工具(如傅里叶变换或谱域滤波),将用户交互序列的特征转换到频域空间处理。设计特定滤波策略,增强表征用户偏好细节的高频特征分量,同时抑制低频冗余噪声或干扰信号,以有效缓解过度平滑问题。

⑦FMLP-Rec:一种隐式序列去噪(Implicit Sequence Denoising,ISD)的推荐模型,旨在通过引入傅里叶变换来增强序列的表示能力,其充分利用频域特性,有效抑制序列中的噪声,从而提高推荐的鲁棒性和准确性。

⑧BSARec:该模型认为自注意力机制具有低通滤波性,提出使用傅里叶变换获取用户兴趣的低频和高频特征。通过高通滤波对高频信息进行强化,缓解序列表示中的过度平滑问题,进一步提升推荐性能。

(5)基于增强视图优化对比学习的SR模型

这类模型旨在结合增强视图优化与对比学习框架,提升推荐性能和鲁棒性。

⑨CaDiRec:一种上下文感知的扩散对比学习模型,用于生成更合理的数据增强视图,以提升SR性能。

⑩MSDCCL:一种软硬去噪策略相结合的模型,可提升在用户历史交互中识别真实兴趣的能力,并用对比学习增强表示的区分性。

⑪FENRec:一种改进的对比学习框架FENRec,用未来多个时间步的行为来构造更细粒度的标签,替代传统的二元标签,并生成在训练后期仍能保持“困难性”的负样本,以增强对比学习的区分能力。

3.4　实验结果与分析

本节通过多个综合实验验证本文模型MCMD-SR的优势及其有效性。在第3.4.1节,将本文模型与11种相关的模型进行实验对比,详细分析并阐明本文模型的优势,以回答RQ1的问题;在第3.4.2节,针对本文模型的4大构件设计了4种变体模型,进行消融实验对比,以回答RQ2;在第3.4.3节,对模型的3个主要关键参数进行调优,以回答RQ3。

3.4.1　对比实验(RQ1)

为评估所提模型MCMD-SR的效果,将各模型在4个不同的经典数据集(Beauty、Yelp、LastFM、ML-1M)上进行对比。我们采用广泛使用的TOP-K指标,HR@K和NDCG@K(在表中分别简写为H@K,N@K)来评估推荐列表,其中K设置为5、10和20,具体结果如表5所示。

由表5可看出,本文模型MCMD-SR在指标HR@K与NDCG@K上均优于对比模型。例如,本文模型MCMD-SR相对于模型caser在4个数据集的HR@K指标上的12个提升比例中,最少达到52.62%。

(1)与基于CNN或RNN的模型对比

①尽管基于CNN的Caser模型是序列推荐领域的早期经典模型,其通过卷积操作捕捉局部序列模式的设计仍具有重要影响。但Caser模型卷积核的有限感受野与局部特征归纳偏好,会导致其对长序列中交互序列的全局依赖关系建模能力不足。相较于Caser模型,本文模型在HR@K和NDCG@K指标上分别最少提升52.62%、71.43%。

②基于RNN的GRU4Rec模型弥补了Caser模型捕捉长期依赖关系的不足,通过其门控机制(重置门与更新门),GRU4Rec模型能保留长期记忆,有效捕捉用户早期行为对当前兴趣的影响。然而,尽管GRU4Rec模型捕捉长距离依赖方面有所改进,但对全局相关性的学习能力不足。相较于GRU4Rec模型,本文模型在HR@K和NDCG@K指标上分别最少提升34.26%、48.44%。

表5 各模型对比结果

Table 5 Comparison of the results of different models

模型	Beauty						Yelp					
	H@5	H@10	H@20	N@5	N@10	N@20	H@5	H@10	H@20	N@5	N@10	N@20
Caser	0.013 0	0.025 0	0.040 3	0.008 2	0.012 1	0.015 9	0.010 4	0.017 9	0.031 7	0.006 3	0.008 7	0.012 1
GRU4Rec	0.018 1	0.029 2	0.047 4	0.011 3	0.014 9	0.019 5	0.013 1	0.024 0	0.041 1	0.007 9	0.011 3	0.015 6
SASRec	0.035 0	0.056 3	0.085 7	0.022 2	0.029 1	0.036 4	0.014 0	0.024 1	0.040 1	0.009 2	0.012 4	0.016 4
BERT4Rec	0.045 1	0.070 4	0.104 7	0.029 6	0.037 7	0.046 3	0.025 9	0.041 9	0.069 9	0.016 4	0.021 6	0.028 6
DuoRec	0.067 3	0.094 5	0.128 8	0.048 2	0.056 9	0.065 5	0.026 2	0.043 7	0.070 6	0.016 3	0.022 0	0.028 7
FEARec	0.067 3	0.093 7	0.129 2	0.048 1	0.056 6	0.065 6	0.024 4	0.039 4	0.063 0	0.015 5	0.020 3	0.026 2
FMLP-Rec	0.038 5	0.059 8	0.090 2	0.025 3	0.032 1	0.039 7	0.016 2	0.027 8	0.046 2	0.009 9	0.013 6	0.018 2
BSARec	0.071 0	0.096 9	0.132 1	0.050 7	0.059 0	0.067 9	0.024 6	0.041 8	0.069 2	0.015 4	0.020 9	0.027 8
CaDiRec	0.047 5	0.071 4	0.101 2	0.031 8	0.039 5	0.047 0	0.022 2	0.036 2	0.060 7	0.013 9	0.018 4	0.024 5
MSDCCL	0.052 2	0.071 4	0.095 5	0.037 8	0.043 9	0.050 0	0.024 3	0.041 8	0.069 5	0.015 1	0.020 7	0.027 7
FENRec	0.069 5	0.097 0	0.131 8	0.049 0	0.057 9	0.066 7	0.027 4	0.045 7	0.074 2	0.017 5	0.023 4	0.030 5
MCMD-SR	**0.072 7**	**0.098 7**	**0.132 5**	**0.052 0**	**0.060 3**	**0.068 8**	**0.030 4**	**0.049 2**	**0.076 6**	**0.018 3**	**0.024 4**	**0.031 2**
Improve/%	**2.39**	**1.75**	**0.30**	**2.56**	**2.20**	**1.33**	**10.95**	**7.66**	**3.23**	**4.57**	**4.27**	**2.30**
模型	LastFM						ML-1M					
	H@5	H@10	H@20	N@5	N@10	N@20	H@5	H@10	H@20	N@5	N@10	N@20
Caser	0.022 9	0.033 9	0.056 9	0.015 9	0.019 3	0.025 0	0.105 1	0.168 5	0.257 3	0.065 7	0.086 0	0.108 5
GRU4Rec	0.029 4	0.044 0	0.060 6	0.021 1	0.025 6	0.029 8	0.123 8	0.197 4	0.292 5	0.077 6	0.101 2	0.125 3
SASRec	0.048 6	0.072 5	0.098 2	0.036 8	0.044 4	0.050 8	0.133 6	0.213 9	0.325 7	0.086 2	0.112 0	0.140 0
BERT4Rec	0.028 4	0.044 0	0.077 1	0.018 2	0.023 1	0.031 4	0.145 2	0.219 0	0.325 5	0.095 3	0.119 1	0.146 0
DuoRec	0.041 3	0.057 8	0.082 6	0.028 9	0.034 1	0.040 3	0.182 3	0.270 7	0.372 2	0.122 9	0.151 3	0.176 9
FEARec	0.026 6	0.047 7	0.056 9	0.019 5	0.026 4	0.028 7	0.187 7	0.271 5	0.371 2	0.123 8	0.150 8	0.176 1
FMLP-Rec	0.027 5	0.042 2	0.053 2	0.020 3	0.024 9	0.027 7	0.135 4	0.214 9	0.329 3	0.086 4	0.112 0	0.140 9
BSARec	0.048 6	0.067 0	0.106 4	0.034 9	0.040 6	0.050 4	0.184 9	0.269 7	0.379 5	0.124 9	0.152 4	0.180 0
CaDiRec	0.048 8	0.073 2	0.106 4	0.034 1	0.042 2	0.050 5	0.127 0	0.201 7	0.293 0	0.084 4	0.108 5	0.131 3
MSDCCL	0.045 5	0.052 0	0.087 1	0.028 8	0.030 8	0.039 4	0.135 1	0.209 5	0.306 2	0.087 3	0.111 2	0.135 6
FENRec	0.042 2	0.066 1	0.094 5	0.028 8	0.036 4	0.043 5	0.189 9	0.273 3	0.384 1	0.128 1	0.155 0	0.183 0
MCMD-SR	**0.051 4**	**0.075 2**	**0.111 9**	**0.037 4**	**0.045 1**	**0.054 2**	**0.194 4**	**0.282 3**	**0.392 7**	**0.129 9**	**0.158 3**	**0.186 0**
Improve/%	**5.33**	**2.73**	**5.17**	**1.63**	**1.58**	**6.69**	**2.37**	**3.29**	**2.24**	**1.41**	**2.13**	**1.64**

注：本文模型MCMD-SR的实验数据以加粗字体标记，为便于比较，利用下划线来突显对比模型中表现最佳的数据，最后一行给出本文模型相对于某一最佳对比模型的性能提升情况（粗体表示）。

基于CNN或RNN的模型能有效捕捉交互序列中的短期或长期依赖关系以进行SR。然而，受限于其串行的循环结构，使其无法计算序列中任意两个物品间的全局相关性。本文提出的MCMD-SR模型利用自注意力掩码机制，旨在学习具有区分度的全局项目表征，从而有效提升推荐多样性。

（2）与基于自注意力机制的SR模型对比

①SASRec模型作为基于Transformer自注意力机制的SR模型，通过多头注意力机制对交互序列中的长距离依赖关系进行建模，但其在处理深层网络时易产生过度平滑现象导致长序列表征退化，进而限制推荐性能的进一步提升。相较于SASRec模型，本文模型在HR@K和NDCG@K指标上分别最少提升3.72%、1.63%。②BERT4Rec模型是一种基于双向自注意力机制的模型，能充分利用上下文信息，从而弥补SASRec模型中信息流单向传递的不足。然而，由于从双向获取信息，BERT4Rec模型在一定程度上加剧了过度平滑现象。相较于BERT4Rec模型，本文模型在HR@K和NDCG@K指标上分别最少提升9.59%、9.09%。

基于自注意力机制的SR模型能捕捉更远距离的依赖关系和更全面的上下文信息，然而，这类模型通常也会加剧过度平滑问题，导致不同用户兴趣表示趋同。即，仅通过自注意力掩码不足以充分缓解过度平滑问题。本文提出的MCMD-SR模型采用自注意力掩

码机制结合对比学习共同缓解过度平滑问题,进一步提升推荐效果。

(3)与基于对比学习及特征增强的SR模型对比

①DuoRec模型创新性地采用双分支架构,融合用户的多兴趣表示学习与对比学习。相较于DuoRec模型,本文模型在HR@K和NDCG@K指标上分别最少提升2.87%、4.63%。

②FEARec利用多视角特征增强与行为语义建模来弥补传统SR模型语义理解薄弱的问题。相较于FEARec模型,本文模型在HR@K和NDCG@K指标上分别最少提升2.55%、4.88%。

基于对比学习及特征增强的SR模型通过对比学习和特征增强来提升模型性能。然而,随着信息的不断融合和传播,这些模型可能会导致用户与项目交互序列中的个性化特征丢失。本文提出的MCMD-SR模型通过特征增强与对比学习相结合的方法,旨在保留用户与项目交互序列的个性化特征。

(4)与基于频域建模缓解过度平滑的SR模型对比

①FMLP-Rec模型引入频域滤波、局部感受野控制以及残差结构等手段,有效保持不同位置序列表示之间的差异性,防止信息的过度融合,这能在一定程度上缓解过度平滑问题,进一步提升模型的表达能力和推荐性能。相较于FMLP-Rec模型,本文模型在HR@K和NDCG@K指标上分别最少提升19.25%、32.01%。

②BSARec模型认为自注意力机制具有低通滤波性,会导致过度平滑问题。为缓解这一问题,BSARec模型引入傅里叶变换调整用户兴趣的高低频信号,从而有效缓解过度平滑问题。相较于BSARec模型,本文模型在HR@K和NDCG@K指标上分别最少提升0.30%、1.33%。

基于频域建模缓解过度平滑的SR模型,通常通过调整频域信息来避免信息的过度融合,从而减轻序列推荐中的过度平滑问题。然而,在特征提取过程中,高频和低频特征往往存在重叠,导致相同特征被重复编码,从而引发特征冗余。我们不仅关注过度平滑的缓解,而且考虑特征冗余带来的负面影响。本文提出的MCMD-SR模型结合自注意力掩码机制和对比学习机制来缓解过度平滑问题,同时引入多维度自适应去过度相关机制,降低特征维度中的冗余信息,以提高序列推荐的整体性能。

(5)与基于增强视图优化对比学习的SR模型对比

①CaDiRec模型利用上下文感知扩散模型生成语义一致的增强视图,从而避免随机增强过程中可能引发的语义漂移问题。相较于CaDiRec模型,本文模型在HR@K和NDCG@K指标上分别最少提升2.73%、3.17%。

②MSDCCL模型通过结合软硬去噪策略与对比学习框架,有效减少噪声对序列表示的干扰,从而提升模型的鲁棒性和推荐性能。相较于MSDCCL模型,本文模型在HR@K和NDCG@K指标上分别最少提升10.22%、12.64%。

③FENRec模型引入未来交互的软标签以及持久性的困难负样本,有效提升用户历史信息的利用率,增强模型的表征能力。相较于FENRec模型,本文模型在HR@K和NDCG@K指标上分别最少提升0.53%、1.41%。

基于增强视图优化对比学习的SR模型中,CaDiRec模型的训练成本较高,且生成质量依赖于扩散模型的建模能力;MSDCCL模型在噪声识别与剔除方面依赖于先验规则或策略设计,难以适应用户行为模式的多样性;而FENRec模型主要聚焦于缓解数据稀疏性,但在局部行为细节的建模上仍显不足。本文提出的MCMD-SR模型在保持序列结构语义的同时,利用自注意力掩码机制生成新的特征表示,有效避免高成本的生成过程;与MSDCCL模型相比,所提模型中引入的去相关机制能够有效抑制冗余特征,提升特征表示的多样性与泛化能力;同时,区别于依赖未来标签的FENRec模型,所提模型无需引入未来信息,即可捕捉序列中的关键兴趣动态,从而实现更高的推荐精度与鲁棒性。

综上分析,本文模型MCMD-SR在4个数据集Beauty、Yelp、LastFM、ML-1M上与最优的对比模型相比,HR@K值最少提升0.3%、3.23%、2.73%、2.24%,平均最少提升2.13%;NDCG@K值最少提升1.33%、2.30%、1.63%、1.41%,平均最少提升1.67%。从对比结果来看,MCMD-SR模型在Yelp和LastFM数据集上提升最大,这是由于这2个数据集均存在较多具有相似交互序列的现象,所提模型可极大地降低相似特征的传播,实现更佳的推荐效果;在数据集Beauty和ML-1M上提升较大,原因在于这2个数据集相似交互的情况仍存在但相对较少,所提模型可在一定程度上降低相似特征的传播,实现较好的推荐效果。可见,本文模型在用户交互序列过度相似的数据集下有更好的表现。

为验证本文所提方法在性能提升方面的统计显著性,我们将其与各基线模型进行配对t检验(paired t-test),实验结果如表6所示。

在保持模型结构与超参数一致的条件下,仅改变随机种子对所有模型进行10次独立重复实验,并记录每次运行的NDCG@10值。之所以选择NDCG@10作为统计检验指标,是因为该指标能够同时衡量命中情况与排序质量,相较于HR更能反映序列推荐任务的效果;同时,Top-10是实际系统中最常采用且具有代表性的推荐列表长度;此外,NDCG的取值为连续

表6 MCMD-SR模型与各对比模型在4个数据集上的NDCG@10指标值配对t检验结果

Table 6 Paired t-test results of NDCG@10 values between MCMD-SR and each baseline on four datasets

模型	Beauty	Yelp	LastFM	ML-1M
Caser	1.14×10^{-17}	2.57×10^{-13}	1.07×10^{-13}	1.49×10^{-18}
GRU4Rec	1.25×10^{-17}	1.62×10^{-11}	7.50×10^{-13}	1.53×10^{-17}
SASRec	3.19×10^{-15}	9.50×10^{-11}	8.19×10^{-3}	2.01×10^{-17}
BERT4Rec	8.22×10^{-16}	8.81×10^{-8}	1.76×10^{-14}	6.43×10^{-16}
DuoRec	5.18×10^{-7}	1.06×10^{-6}	3.40×10^{-13}	5.95×10^{-9}
FEARec	1.10×10^{-6}	8.13×10^{-8}	5.28×10^{-13}	4.78×10^{-11}
FMLP-Rec	2.13×10^{-17}	1.52×10^{-15}	6.86×10^{-14}	1.24×10^{-16}
BSARec	1.50×10^{-3}	3.61×10^{-7}	4.06×10^{-7}	2.64×10^{-8}
CaDiRec	9.08×10^{-14}	5.85×10^{-8}	5.17×10^{-6}	3.54×10^{-16}
MSDCCL	1.69×10^{-12}	5.84×10^{-7}	1.59×10^{-10}	8.30×10^{-17}
FENRec	6.61×10^{-5}	7.58×10^{-3}	1.61×10^{-9}	3.63×10^{-5}

型且方差稳定，更符合配对t检验对数据分布的基本要求。因此，基于NDCG@10的显著性分析较为合理。随后，我们根据10组实验得到的NDCG@10值，分别对本文方法与各基线模型进行配对t检验。当p-value小于0.05时，说明性能差异具有统计显著性。从NDCG@10指标值的配对t检验结果来看，本文方法相较所有基线模型的提升都具有统计显著性，这表明所提方法具有稳定且可靠的增强效果。

3.4.2 消融实验(RQ2)

为验证自注意力掩码机制、对比学习模块、列间自适应去过度相关、层间自适应去过度相关这4个构件对序列推荐模型性能的影响，设计了4种变体模型，分别是Ours-1、Ours-2、Ours-3、Ours-4并将它们与本文模型进行对比，以证明各构件存在的必要性。

①Ours-1不采用自注意力掩码机制构件，而是替换成随机掩码机制。然后，将得到的随机掩码矩阵用于对比学习，并进行列间、层间自适应去过度相关操作。将该变体模型用于对比，主要验证自注意力掩码机制对降低交互序列特征相似性的有效性，同时考察在缺乏自注意力掩码机制构件时其他构件的协同效果。

②Ours-2去除对比学习模块构件，未引入其他替代方式，直接将自注意力掩码机制构件、列间自适应去过度相关构件、层间自适应去过度相关构件进行联合优化学习。将该变体模型用于对比，主要阐述对比学习模块构件强化个性化特征的作用，同时考察其他三个构件联合优化学习的有效性。

③Ours-3不引入列间自适应去过度相关构件，而仅基于层间自适应去过度相关构件、自注意力掩码机制构件以及对比学习模块联合优化学习后进行预测。将该变体模型用于对比，主要评估列间自适应去过度相关构件在降低列间特征冗余方面的能力，同时考察其他三个构件的相互作用。

④Ours-4不使用层间自适应去过度相关构件，而仅基于列间自适应去过度相关构件、自注意力掩码机制构件以及对比学模块构件共同作用下的推荐效果。将该变体模型用于对比，主要评估层间自适应去过度相关降低层间特征冗余的能力，同时考察其他三个构件的相互作用。

需说明的是，仅使用任意一个构件和四个构件均不使用的变体模型因效果不如其他模型，故无需讨论；另外，对使用任意两个构件的变体模型(Ours-5、Ours-6、Ours-7、Ours-8、Ours-9、Ours-10)，也进行两两组合的消融实验。变体模型组件构成情况如表7所示。

表7 变体模型组件构成情况

Table 7 Component composition of the variant models

变体模型	自注意掩码机制	对比学习模块	列间自适应去过度相关	层间自适应去过度相关
Ours-1	×	√	√	√
Ours-2	√	×	√	√
Ours-3	√	√	×	√
Ours-4	√	√	√	×
Ours-5	√	√	×	×
Ours-6	√	×	√	×
Ours-7	√	×	×	√
Ours-8	×	√	√	×
Ours-9	×	√	×	√
Ours-10	×	×	√	√
MCMD-SR	√	√	√	√

在Beauty、Yelp、LastFM、ML-1M这4种公开数据集上进行消融实验，考虑到变体模型的目的是验证模型构件对推荐性能的影响，这里选择HR@10和NGCD@10(表中简称H@10和N@10)作为评价指标，以反映推荐性能，如表8所示。

从表8可看出，自注意力掩码机制构件对推荐效果影响最大，使用随机注意力掩码的变体模型Ours-1在HR@10和NDCG@10这2个指标上表现不如包含自注意力掩码机制构件的模型；列间去过度相关构件对推荐效果影响很大，不考虑列间自适应去过度相关构件的变体模型Ours-3在这2个指标上表现不如引入列间自适应去过度相关构件的模型；对比学习构件对推荐效果影响较大，不考虑对比学习的变体模型Ours-2在这2个指标上表现不如使用对比学习构件的模型；层间去过度相关构件对推荐效果有一定影响，不考虑层间自适应去过度相关的变体模型Ours-4

表8　消融实验结果对比

Table 8　Comparison of the results of ablation study

模型	Beauty		Yelp		LastFM		ML-1M	
	H@10	N@10	H@10	N@10	H@10	N@10	H@10	N@10
Ours-1	0.072 7	0.041 7	0.032 6	0.016 3	0.062 4	0.036 3	0.272 0	0.152 1
Ours-2	0.096 2	0.058 7	0.037 7	0.018 3	0.074 3	0.042 3	0.273 5	0.152 9
Ours-3	0.092 7	0.057 0	0.039 3	0.019 8	0.067 9	0.040 6	0.271 5	0.152 0
Ours-4	0.098 6	0.058 5	0.041 4	0.020 7	0.071 6	0.043 0	0.273 2	0.155 1
Ours-5	0.092 8	0.055 1	0.031 4	0.016 0	0.064 1	0.038 3	0.262 2	0.148 5
Ours-6	0.095 9	0.056 6	0.029 8	0.014 4	0.070 5	0.040 1	0.264 6	0.149 0
Ours-7	0.090 1	0.055 0	0.027 5	0.013 5	0.067 2	0.037 5	0.262 5	0.146 5
Ours-8	0.072 4	0.039 6	0.024 7	0.012 8	0.058 7	0.034 0	0.263 0	0.148 5
Ours-9	0.066 5	0.038 3	0.022 8	0.011 8	0.055 0	0.031 6	0.260 8	0.145 7
Ours-10	0.069 6	0.039 8	0.021 0	0.010 3	0.061 5	0.033 0	0.262 9	0.146 9
MCMD-SR	**0.098 7**	**0.060 3**	**0.049 2**	**0.024 4**	**0.075 2**	**0.045 1**	**0.282 3**	**0.158 3**
Improve/%	**0.10**	**2.65**	**15.85**	**15.16**	**1.20**	**4.66**	**3.12**	**2.02**

注:本文模型MCMD-SR的实验数据以加粗字体表示,为便于比较,利用下划线来突显变体模型中表现最佳的数据,表底给出MCMD-SR模型相对于某一最佳变体模型的性能提升情况(粗体表示)。

在这2个指标上表现不如引入该构件的模型。

(1)从结合对比学习模块构件、列间自适应去过度相关构件以及层间自适应去过度相关构件的变体模型(Ours-1)来看,自注意力掩码机制构件对推荐效果影响最大。虽然对特征矩阵使用随机掩码能屏蔽一些相似特征,但是不引入自注意力掩码机制构件,缺乏对用户项目交互序列特征重要性的有效区分,导致一些重要特征被丢失。从表8可看出,与Ours-1相比,保留了自注意力掩码机制构件的所提模型MCMD-SR性能显著提升,在4个数据集中,HR@10和NDCG@10指标分别平均提升20.19%和21.87%,其中,在HR@10指标中至少提升3.65%,最多提升可达33.74%。可见,自注意力掩码机制构件对模型性能提升具有重要贡献。

(2)从结合自注意力掩码机制构件、列间自适应去过度相关构件以及层间自适应去过度相关构件的变体模型(Ours-2)来看,对比学习模块构件对推荐效果影响较大。尽管自注意力掩码衰减机制,能引导模型学习差异化的特征表示,然而如果不采用对比学习模块构件,个性化特征仍不够显著。从表8可看出,与Ours-2相比,保留了对比学习模块构件的所提模型MCMD-SR性能有一定提升,在4个数据集中,HR@10和NDCG@10指标平均提升7.56%和9.32%,其中,在HR@10指标中至少提升1.20%,最多提升可达23.37%。可见,对比学习模块构件能较大地增强模型的推荐性能。

(3)从结合自注意力掩码机制构件、对比学习模块构件以及层间自适应去过度相关构件的变体模型(Ours-3)来看,列间自适应去过度相关构件对推荐效果影响很大。虽然通过自注意力掩码机制可缓解过度平滑现象,但是不引入列间自适应去过度相关构件,无法降低列间特征相关,导致局部冗余信息被传播。从表8可看出,与Ours-3相比,保留了列间自适应去过度相关构件的所提模型MCMD-SR性能大幅提升,在4个数据集中,HR@10和NDCG@10指标分别平均提升9.94%和9.57%,其中,在HR@10指标中至少提升3.83%,最多提升可达20.12%。可见,列间自适应去过度相关构件能降低列间特征相关,减少冗余信息传播,对模型性能提升具有很大作用。

(4)从结合自注意力掩码机制构件、对比学习模块构件以及列间自适应去过度相关构件的变体模型(Ours-4)来看,层间自适应去过度相关构件对推荐效果有一定影响。尽管列间自适应去过度相关构件能降低局部特征相关性,然而不引入层间自适应去过度相关构件,无法降低全局特征相关性,导致全局冗余信息被传播。从表8可看出,与Ours-4相比,保留了层间自适应去过度相关构件的所提模型MCMD-SR性能有所提升,在4个数据集中,HR@10和NDCG@10指标分别平均提升5.99%和6.21%,其中,在HR@10指标中至少提升0.1%,最多提升可达15.85%。可见,层间自适应去过度相关构件对模型的推荐性能具有提升作用。

综上分析可知:①自注意力掩码机制可对过度相似的交互序列特征根据注意力分数进行替换重组,形成更高质量的交互特征,进而缓解过度平滑现象;②列间自适应去过度相关能从特征矩阵的列间维度降低局部特征相关,减少局部冗余信息的传播;③对比学习模块可增强用户-项目交互序列的个性化特征,提高交互序列的嵌入质量;④层间自适应去过度相关

能从特征矩阵的层级间降低全局特征相关,减少全局冗余信息的传播;⑤4个核心构件进行多任务联合优化学习,各构件相互补充、相互促进,共同缓解过度平滑问题,提升序列推荐模型的泛化性和准确性。

此外,为更好地揭示组件之间的相互作用和依赖关系,我们还做了两两组合消融实验(见表8),该实验结果清晰地揭示了两两构件之间的交互机制以及它们在联合优化中的依赖关系。

(1)Ours-6(自注意力掩码机制+列间自适应去过度相关)表现最优,说明"前置差异增强+局部冗余压缩"构成最具协同效应的结构。自注意力掩码机制首先显著减少输入特征间的相似度,使模型获得更具辨识度的表示;随后,列间自适应去过度相关模块在维度层面进一步去除局部冗余特征,使得自注意力掩码机制产生的差异能够被充分保留并有效利用,形成了强依赖关系与最稳定的协同效果。

(2)Ours-5(自注意力掩码机制+对比学习模块)体现了"特征差异制造与特征差异放大"的双向增强模式:自注意力掩码机制为对比学习模块提供了清晰的差异表示,而对比学习模块在此基础上进一步强化表示的区分性。然而,由于缺少去冗余模块以进一步降低特征冗余,其性能略低于Ours-6。

(3)Ours-7(自注意力掩码机制+层间自适应去过度相关)通过层间自适应去过度相关抑制深层网络中产生的全局冗余信息,从而保持自注意力掩码机制生成的差异不在深层传播过程中被再次平滑化,形成"差异保持型"的协同模式。

(4)Ours-8(对比学习模块+列间自适应去过度相关)虽然具备"差异放大+局部去冗余"的互补机制,但由于缺乏自注意力掩码机制提供前置差异,其整体提升幅度受限,表现弱于所有包含自注意力掩码机制的组合。

(5)Ours-10(列间自适应去过度相关+层间自适应去过度相关)仅依赖多维度去冗余模块消除局部与全局特征相关性,尽管两者为正交互补关系,但缺少差异增强方式,性能提升有限。

(6)Ours-9(对比学习模块+层间自适应去过度相关)性能最弱,主要由于缺乏自注意力掩码机制与列间自适应去过度相关模块的前置作用,使后续模块的增益受限。未经掩码处理的高度相似特征输入制约了对比学习模块的判别能力,而层间自适应模块虽能处理相关性,但面对未被降低的源头特征冗余,其优化能力有限。

由此发现,两两组合消融实验表明模型构件间存在明确的功能依赖链:自注意力掩码机制负责前置差异制造,对比学习模块进一步放大差异,列间自适应去过度相关与层间自适应去过度相关模块分别在局部与全局这两个维度减少冗余信息,共同形成层层递进的协同关系。这一互补性结构与依赖机制验证了4个核心构件联合使用的必要性与整体设计的合理性。

3.4.3 参数敏感度实验(RQ3)

本文对模型中的对比学习损失权重λ_1、列间自适应去过度相关损失权重λ_2、层间自适应去过度相关损失权重λ_3三个关键参数进行实验分析,阐述它们在不同数据集上对模型性能的影响。注意,因自注意力掩码机制在多任务联合优化学习中起主导作用,故将其损失权重设为1,而无需单独分析。

(1)对比学习损失权重λ_1的影响

λ_1是控制对比学习辅助任务占联合学习任务比重的参数,若取值不同,则对比学习所占联合任务中的损失权重不同,也说明对比学习对模型有不同程度的影响。为达到较优效果,我们对λ_1进行实验,从集合$\{0.000\,1, 0.000\,5, 0.01, 0.05, 1\}$中选取不同的$\lambda_1$值,观察不同的$\lambda_1$值在不同数据集下HR@10和NDCG@10指标值变化情况,图2所示。

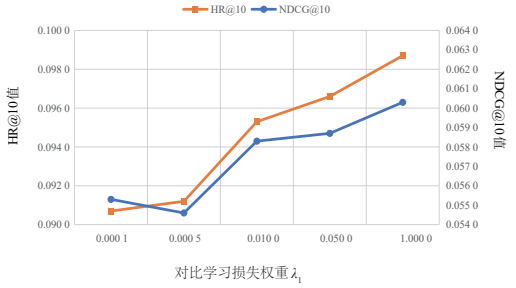
从图2(a)可见,在Beauty数据集上,随着λ_1的增大,模型性能整体呈上升趋势,尤其在$\lambda_1=1$时达到最佳,表明较大的对比学习信号有助于提升表示学习效果。从图2(b)可见,在Yelp数据集上,随着λ_1的增大,模型性能整体上呈上升趋势,虽然在$\lambda_1=0.05$时略微下降,但是性能指标仍然处在较高值,尤其在$\lambda_1=1$时达到最佳,表明较大的对比学习信号能够引导模型学习到更好的表示。从图2(c)可见,在LastFM数据集上,模型对λ_1的变化极为敏感,最优性能出现在$\lambda_1=0.000\,5$时;而当λ_1继续增大时,性能显著下降,说明对比学习信号过强可能会扰乱序列特征的学习。从图2(d)可见,在ML-1M数据集上,$\lambda_1=0.000\,1$时取得最优的性能,过大的λ_1会导致性能下降,说明在大规模稠密数据中,对比损失应适度引导模型而非主导优化。

可见,在各数据集上,参数λ_1最优值取不同,我们认为是合理的。原始数据集的特点不一样,合适的λ_1取值能更好地平衡对比学习任务与其他任务间的关系,从而使模型达到更好的推荐效果。

(2)列间自适应去过度相关损失权重λ_2的影响

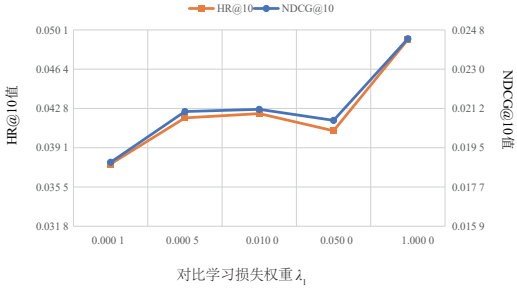
λ_2是控制列间自适应去过度相关辅助任务占联合学习任务比重的参数,若取值不同,则列间自适应去过度相关所占联合任务中的损失权重不同,也说明列间自适应去过度相关对模型有不同程度影响。为达最优效果,我们对λ_2进行实验,从集合$\{0.000\,1, 0.000\,2, 0.01, 0.02, 1\}$中选取不同$\lambda_2$值,观察不同$\lambda_2$值在不同数据集下HR@10和NDCG@10指标值变化情况,如图3所示。

从图3(a)可见,在Beauty数据集上,随着λ_2的增大,模型性能整体呈上升趋势,尤其在$\lambda_2=1$时达到最



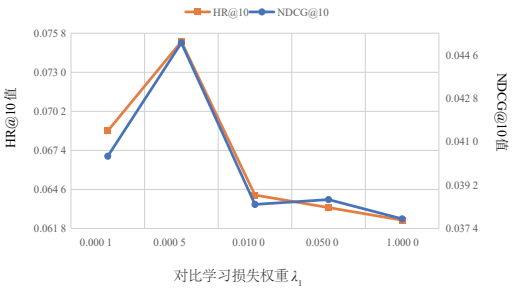
(a) λ_1 在 Beauty 上的效果

(a) The effect of λ_1 on beauty



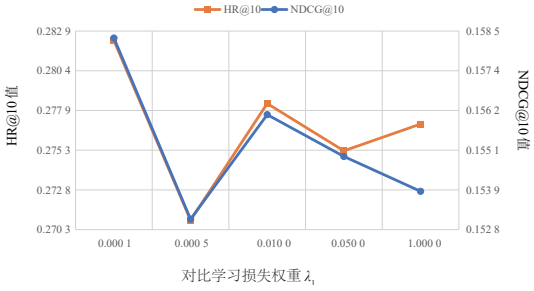
(b) λ_1 在 Yelp 上的效果

(b) The effect of λ_1 on yelp



(c) λ_1 在 LastFM 上的效果

(c) The effect of λ_1 on lastFM

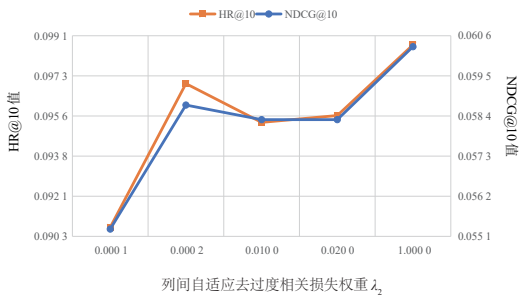


(d) λ_1 在 ML-1M 上的效果

(d) The effect of λ_1 on ML-1M

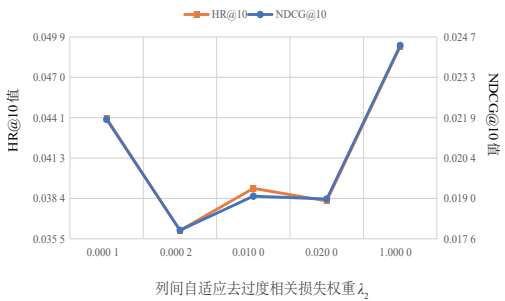
图 2　对比学习损失权重 λ_1 的影响

Figure 2　Effect of the contrastive loss weight λ_1



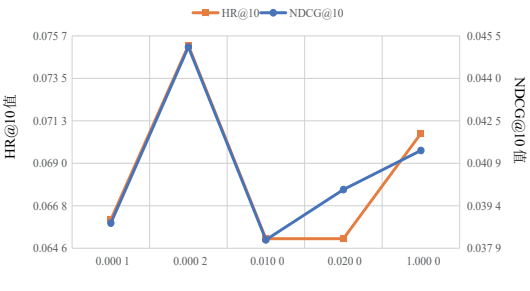
(a) λ_2 在 Beauty 上的效果

(a) The effect of λ_2 on beauty



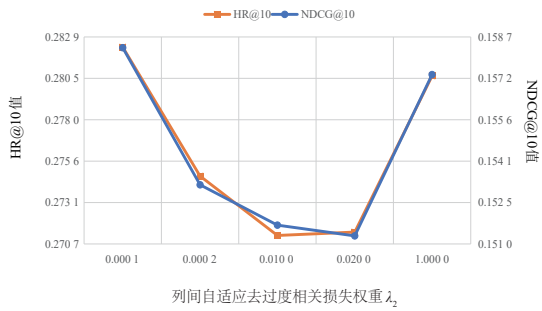
(b) λ_2 在 Yelp 上的效果

(b) The effect of λ_2 on yelp



(c) λ_2 在 LastFM 上的效果

(c) The effect of λ_2 on lastFM



(d) λ_2 在 ML-1M 上的效果

(d) The effect of λ_2 on ML-1M

图 3　列间自适应去过度相关损失权重 λ_2 的影响

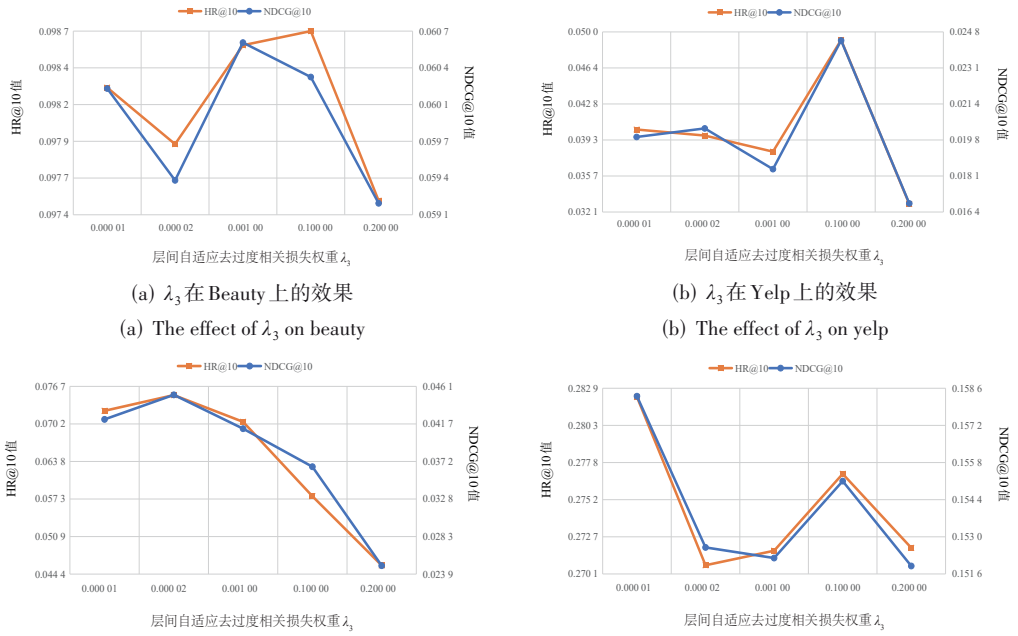
Figure 3　Effect of the column-wise adaptive decorrelation loss weight λ_2

佳。该数据集在列间特征维度中存在一定程度的冗余，较强的列间自适应去过度相关约束可有效提升推荐效果。从图3(b)可见，在Yelp数据集上，当$\lambda_2=0.000\,1$，模型对冗余特征控制不足，性能不理想；当λ_2继续增大至0.000 2时，模型性能显著下降；而在$\lambda_2=0.01$附近时性能略有回升，但仍未达到最优；当$\lambda_2=1$时，性能明显提升，表明强约束在该数据集上能有效缓解特征冗余，改善推荐效果。从图3(c)可见，在LastFM数据集上，模型对λ_2的变化较为敏感，最优性能出现在$\lambda_2=0.000\,2$时，表明适度的列间自适应去过度相关约束有助于提升表示质量；而当λ_2进一步增大时，性能整体上处于较低水平，说明过强的去相关信号干扰了序列信息的有效建模。从图3(d)可见，在ML-1M数据集上，$\lambda_2=0.000\,1$时模型表现最优，说明较小的约束即可带来正面效果；当λ_2继续增大至0.02时，模型性能大幅下降，表明在该稠密数据集上，过强的列间自适应去过度相关约束可能削弱了协同信号，导致特征学习能力降低；从0.02进一步增至$\lambda_2=1$时，模型性能有所回升，但仍未恢复到最佳水平。

可见，在不同数据集上，参数λ_2最优取值不同，我们认为这是合理的。在不同的数据集中，具有不同的局部用户交互序列特征，合适的λ_2取值能更有效地降低不同数据集的冗余度，从而使模型达到更好的推荐效果。

(3)层间自适应去过度相关损失权重λ_3的影响

λ_3是控制层间自适应去过度相关辅助任务占联合学习任务比重的参数，若取值不同，则层间自适应去过度相关所占联合任务中的损失权重不同，也说明层间自适应去过度相关对模型有不同程度的影响。为达较优效果，我们对λ_3值进行实验，从集合$\{0.000\,01, 0.000\,02, 0.001, 0.1, 0.2\}$中选取不同的$\lambda_3$值，观察不同的$\lambda_3$值在不同数据集下HR@10和NDCG@10指标值变化情况，如图4所示。



(a) λ_3在Beauty上的效果

(a) The effect of λ_3 on beauty

(b) λ_3在Yelp上的效果

(b) The effect of λ_3 on yelp

(c) λ_3在LastFM上的效果

(c) The effect of λ_3 on lastFM

(d) λ_3在ML-1M上的效果

(d) The effect of λ_3 on ML-1M

图4 层间自适应去过度相关损失权重λ_3的影响

Figure 4 Effect of the layer-wise adaptive decorrelation loss weight λ_3

从图4(a)可见，在Beauty数据集上，当$\lambda_3=0.001$时，模型的NDCG@10值达到最大，而当$\lambda_3=0.1$时，HR@10值达到最大，说明合适的层间自适应去过度相关权重有助于提升多层表示之间的多样性，从而增强表示能力；当λ_3继续增大时，层间自适应去过度相关信号过强会干扰有效信息的传递，反而降低模型性能。从图4(b)可见，在Yelp数据集上，当$\lambda_3<0.1$时，模型性能较差，表明较弱的层间自适应去过度相关约束不足以缓解全局特征冗余问题；而在$\lambda_3=0.1$时性能显著提升，表明较强的层间自适应去过度相关约束可更有效地抑制无效信息干扰，从而提升模型性能；但当λ_3继续增大到0.2时，模型性能明显下降，说明过强约束可能破坏了有益信息的表达。从图4(c)可见，在LastFM数据集上，模型$\lambda_3=0.000\,02$时达到性能峰值；此后随着λ_3增大，模型性能迅速下降，反映出过强的层间自适应去过度相关信号可能扰乱个性化偏好建模，影响模型的序列学习效果。从图4(d)可见，在ML-1M数据集上，最优性能出现在$\lambda_3=0.000\,01$时，进

一步增大λ_3会造成明显性能下降,说明该稠密数据集对强去相关约束较为敏感;尽管在$\lambda_3=0.1$时性能略有回升,但依然未能恢复到最优效果,这表明在协同信号丰富的数据环境下,较小的去相关约束更有利于提高模型的学习能力。

可见,在不同的数据集上,参数λ_3最优值取不同,我们认为这是合理的。不同的数据集具有不同的全局用户交互序列特征,合适的λ_3取值能更好地平衡联合学习中与列间自适应去过度相关的关系,从而使模型达到更佳的推荐效果。

3.4.4　模型复杂度分析(RQ4)

与对比模型相比,本文模型MCMD-SR引入自注意力掩码机制,根据注意力分数进行特征掩码并使用对比学习突出个性特征,从而缓解了过度平滑问题,又利用多维度自适应去过度相关,从多个维度去除冗余特征,从而提高模型的泛化能力,由于多个模块的增加,模型的时间复杂度有所增加,但更多时间的牺牲换来的推荐性能提升是值得的,具体分析如下。

(1)本文模型MCMD-SR的特征编码利用了BSARec模型,该模型的时间复杂度大致为$O(LN^2D+LN\log N+LND^2)$,L为模型层数,N为序列长度,D为嵌入维度;自注意力掩码机制模块时间复杂度大致为$O(N^2D)$;对比学习模块时间复杂度大致为$O(N^2D)$;多维度自适应去过度相关模块时间复杂度大致为$O(N^2)$。因此,本文模型的时间复杂度大致为$O(LN^2D+LN\log N+LND^2)$,约简为$O(LND^2)$。

(2)就对比模型GRU4Rec(相对较经典的基于改进RNN的序列推荐)的时间复杂度来说,它的时间主要消耗在计算重置门和更新门上。因此,时间复杂度大致为$O(ND)$。

(3)就对比模型BERT4Rec(基于自注意力机制的SR模型中整体表现最佳)的时间复杂度来说,它的时间主要消耗在Transformer编码器这个核心部分,对于每一层编码器,自注意力的计算复杂度大致为$O(N^2D)$,并且通常有多层Transformer编码器。因此,时间复杂度大致为$O(FN^2D)$,F为编码器层数。

(4)就对比模型DuoRec(基于对比学习的SR模型中整体表现最佳)的时间复杂度来说,它的核心是Transformer编码器和对比学习,Transformer编码器部分的时间复杂度大致为$O(FN^2D)$,对比学习模块的时间复杂度大致为$O(N^2D)$。因此,时间复杂度大致为$O(FN^2D+N^2D)$,约简为$O(FND^2)$

(5)就对比模型FMLP-Rec(运行时长相对较短)的时间复杂度来说,它的时间主要消耗在过滤层即傅里叶变换和逆傅里叶变换,过滤层的时间复杂度大致为$O(N\log N)$,并且包含多个学习过滤增强的MLP(Multi-Layer Perceptron)块。因此,时间复杂度大致为$O(LN\log N)$。

(6)就对比模型CaDiRec(相对较新的序列推荐)的时间复杂度来说,它的时间主要消耗在以下几个部分,Transformer部分的时间复杂度为$O(FN^2D)$,Diffusion模型部分的为$O(TN^2D)$,T为扩散步,正负样本对构造部分的为$O(N^2D)$。因此,复杂度大致为$O(FN^2D+TN^2D)$,约简为$O(TND^2)$。

综上分析,各模型的时间复杂度大小关系为:CaDiRec > MCMD-SR > DuoRec = BERT4Rec > GRU4Rec > FMLP-Rec。从时间复杂度上来看,所提模型MCMD-SR并未达到最高,较对比模型CaDiRec更低,尽管略高于其他部分对比模型,但就推荐性能而言,本文模型MCMD-SR最优(见表5),故其设计思想合理且有意义。

所提模型MCMD-SR在推荐性能上虽取得显著优势,但其时间复杂度高于DuoRec、BERT4Rec、GRU4Rec和FMLP-Rec。为验证模型MCMD-SR在大规模数据集场景下的效率与部署可行性,我们将其从推理时间、推理显存以及推荐效果等方面与基线模型进行实验对比。

本文选用最新的Yelp2022数据集(经序列化处理后,用户数为69 428,项目数为41 260,交互次数为1 666 296,平均长度为24,稀疏度为0.999 4)展开实验,该数据集非常贴近大规模数据场景中的高稀疏性、高复杂度推荐环境,能够有效地检验模型在大规模数据下的推理效率与部署可行性。此外,实验采用Top-10作为推荐列表长度,这是推荐模型中最常用且具有广泛代表性的评估设定。实验结果如表9所示。

表9　Yelp2022数据集上模型效率、部署可行性与推荐效果对比

Table 9　Comparison of model efficiency, deployability, and recommendation effect on Yelp2022

模型	推理时间/s	推理内存/GB	HR@10	NDCG@10
CaDiRec	5.95	0.14	0.040 5	0.020 2
DuoRec	1.20	0.12	0.058 5	0.029 7
BERT4Rec	0.86	0.12	0.051 9	0.026 0
GRU4Rec	0.67	0.11	0.032 1	0.016 0
FMLP-Rec	0.81	0.11	0.032 7	0.016 2
MCMD-SR	**1.94**	**0.14**	**0.059 9**	**0.030 3**

注:本文模型MCMD-SR的实验数据以加粗字体标记。

从表9可看出:

(1)虽然所提模型MCMD-SR的时间复杂度理论上高于DuoRec、BERT4Rec、GRU4Rec和FMLP-Rec,但是在实验中,其推理时间仅为1.94 s。相比之下,另一对比模型CaDiRec的推理时间高达5.95 s。这表明MCMD-SR在交互建模过程中并未带来过度的计算

延迟,推理时间依旧保持在可部署范围内,说明模型的复杂度与效率在实际运行中达到良好的平衡。

(2)所提模型MCMD-SR推理显存占用仅为0.14 GB,仅分别比DuoRec、BERT4Rec、GRU4Rec和FMLP-Rec略高0.02 GB、0.02 GB、0.03 GB、0.03 GB,资源开销低,处于合理范围内。提出的多维度去过度相关模块主要利用PCC在列间与层间统计相关性,未引入额外大规模参数,因此显存增加极为有限。

(3)在额外开销稍有增加的前提下,所提模型MCMD-SR获得显著的性能优势(HR@10提升2.4%~86.6%,NDCG@10提升2.0%~89.4%)。这说明引入的自注意力掩码机制、对比学习模块与自适应去过度相关模块在复杂度可接受的前提下带来了可观增益,权衡了效率与性能。

(4)若需进一步部署在超大规模场景,模型也支持以下轻量化方向:

①可调整自注意力掩码机制模块的层数,根据实际需要减少自注意力掩码机制模块层数以降低显存及计算量。

②可在推理阶段关闭对比学习模块,仅保留自注意力掩码机制模块与列间自适应去过度相关模块以降低显存及计算量。

③可关闭层间自适应去过度相关模块,仅保留列间自适应去过度相关模块以降低显存及计算量。

可见,所提模型MCMD-SR在保持显著性能优势的同时,其推理时间与显存占用也能做到与主流序列推荐模型相当,进一步表明所提模型在大规模数据集场景下仍具有较好的推理效率与部署可行性,同时也支持轻量化方式。

4 结束语

本文提出一种MCMD-SR模型,利用自注意力机制计算的分数对特征掩码,减少项目特征间的相似性,提高模型对差异项目的辨识性;利用对比学习,通过最大化正样本相似性和最小化负样本相似性提升嵌入表示的个性化。特征掩码与对比学习共同作用缓解基于注意力机制的SR中的过度平滑问题。利用多维度自适应去过度相关,从列间相似性和层间相似性两个维度共同降低高维特征的相关性,提升模型对关键特征的关注度,再与缓解过度平滑模块相协作,提升模型的泛化能力。

(1)为缓解基于注意力机制的SR中的过度平滑现象,设计自注意力掩码机制结合对比学习模块。通过构建自注意力特征掩码矩阵,以减少项目特征间的相似性,进而提高模型对差异项目特征的辨识性;同时,利用对比学习使模型获得更多个性化特征,并与自注意力掩码机制共同缓解过度平滑问题。

(2)为降低特征过度相关性,设计多维度自适应去过度相关模块。构建一个列间和层间自适应去过度相关模块,从局部和全局降低项目特征维度的过度相关,减少特征冗余度,提升模型对关键特征的关注度和对项目的泛化性。

(3)为验证所提模型效果,展开多个综合实验分析。无论是与经典模型还是最新模型对比,所提模型MCMD-SR在11个对比模型的4个数据集中,HR@K和NDCG@K(K为5、10和20)推荐指标均有明显提升。同时,通过消融实验,验证了模型各构件的必要性。此外,参数敏感度实验优化了模型的关键参数。

所提模型MCMD-SR推荐效果虽更优,但其研究重点在SR,未兼顾除SR以外的推荐任务在缓解过度平滑问题与提高模型泛化能力方面的情形。下一步工作中,我们将积极探索能迁移到SR任务外缓解过度平滑现象并增强模型泛化性的方法。

参考文献

[1] Shin Y, Choi J, Wi H, et al. An attentive inductive bias for sequential recommendation beyond the self-attention[C]// Proceedings of the 38th AAAI Conference on Artificial Intelligence. Vancouver: AAAI Press, 2024: 8984-8992.

[2] Fan Ziwei, Liu Zhiwei, Peng Hao, et al. Addressing the rank degeneration in sequential recommendation via singular spectrum smoothing[C]//Proceedings of the ACM Conference on Recommender Systems. Woodstock: ACM, 2023: 11986.

[3] Zhou Kun, Yu Hui, Zhao W X, et al. Filter-enhanced MLP is all you need for sequential recommendation[C]//Proceedings of the ACM Web Conference 2022. Lyon: ACM, 2022: 2388-2399.

[4] Du Xinyu, Yuan Huanhuan, Zhao Pengpeng, et al. Frequency enhanced hybrid attention network for sequential recommendation[C]//Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval. Taipei, China: ACM, 2023: 78-88.

[5] Cui Ziqiang, Wu Haolun, He Bowei, et al. Context matters: Enhancing sequential recommendation with context-aware diffusion-based contrastive learning[C]//Proceedings of the 33rd ACM International Conference on Information and Knowledge Management. Boise: ACM, 2024: 404-414.

[6] Zhu Xiaofei, Li Liang, Liu Weidong, et al. Multi-level sequence denoising with cross-signal contrastive learning for sequential recommendation[J]. Neural Networks, 2024, 179(6): 106480.

[7] Huang Y H, Lo L, Xie Hongxia, et al. Future sight and tough fights: Revolutionizing sequential recommendation with FENRec[C]//Proceedings of the 39th AAAI Conference on Artificial Intelligence. Philadelphia: AAAI, 2025: 11826-11834.

[8] 钱忠胜, 黄恒, 万子珑. 融合自注意力机制的多行为图对比学习推荐方法[J]. 电子学报, 2024, 52(11): 3684-3698.

Qian Zhongsheng, Huang Heng, Wan Zilong. The multi-behavior graph contrastive learning recommendation method with self-attention mechanism[J]. Acta Electronica Sinica, 2024, 52(11): 3684-3698. (in Chinese)

[9] Luo Yanchen, Li Sihang, Sui Yongduo, et al. Masked graph modeling with multi- view contrast[C]//Proceedings of 2024 IEEE 40th International Conference on Data Engineering. Utrecht: IEEE, 2024: 2584-2597.

[10] Liu Chuang, Wang Yuyao, Zhan Yibing, et al. Where to mask: Structure-guided masking for graph masked autoencoders[C]// Proceedings of the 33rd International Joint Conference on Artificial Intelligence. Jeju: ijcai.org, 2024: 2180-2188.

[11] Fang Taoran, Xiao Zhiqiang, Wang Chunping, et al. DropMessage: Unifying random dropping for graph neural networks[C]// Proceedings of the 37th AAAI Conference on Artificial Intelligence. Washington: AAAI Press, 2023: 4267-4275.

[12] Liu Meng, Gao Hongyang, Ji Suiwang. Towards deeper graph neural networks[C]//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York: ACM, 2020: 338-348.

[13] 刘昕悦, 尹海莲, 臧亚磊, 等. 基于图插值和可变形卷积网络的序列推荐[J]. 计算机研究与发展, 2025, 62(10): 2583-2594.
Liu Xinyue, Yin Hailian, Zang Yalei, et al. A graph-based interpolation sequential recommender with deformable convolutional network[J]. Journal of Computer Research and Development, 2025, 62(10): 2583-2594. (in Chinese)

[14] You Di, Lee K. Context-aware diffusion-based sequential recommendation[C]//Proceedings of 2024 IEEE International Conference on Big Data. Washington: IEEE, 2024: 670-679.

[15] He Zhankui, Zhao Handong, Wang Zhaowen, et al. Query-aware sequential recommendation[C]//Proceedings of the 31st ACM International Conference on Information & Knowledge Management. Atlanta: ACM, 2022: 4019-4023.

[16] Zhang Hao, Cheng Mingyue, Liu Zhiding, et al. Towards automatic sampling of user behaviors for sequential recommender systems[C]//Proceedings of the 34th International Joint Conference on Artificial Intelligence. Montreal: ijcai.org, 2025: 3624-3632.

[17] Pakdaman F, Gabbouj M. Channel-wise feature decorrelation for enhanced learned image compression[J]. IEEE Signal Processing Letters, 2024, 31: 1635-1639.

[18] Hua Tianyu, Wang Wenxiao, Xue Zihui, et al. On feature decorrelation in self-supervised learning[C]//Proceedings of 2021 IEEE/CVF International Conference on Computer vision. Montreal: IEEE, 2021: 9578-9588.

[19] 郭向星, 周魏, 杨正益, 等. 基于自监督图卷积和注意力机制实现隐式反馈降噪的社交推荐[J]. 电子学报, 2025, 53(1): 151-162.
Guo Xiangxing, Zhou Wei, Yang Zhengyi, et al. Denoising implicit feedback with self-supervised graph convolution network and attention mechanism for social recommendation[J]. Acta Electronica Sinica, 2025, 53(1): 151-162. (in Chinese)

[20] Zhu Qiuyu, Wang Hao, Zu Xuewen, et al. Multi-stage feature decorrelation constraints for improving CNN classification performance[C]//Proceedings of 2023 China Automation Congress. Chongqing: IEEE, 2023: 9219-9224.

[21] Hidasi B, Karatzoglou A, Baltrunas L, et al. Session-based recommendations with recurrent neural networks[PP/OL]. V4. arXiv (2016-03-29)[2025-06-20]. https://arxiv.org/abs/1511.06939.

[22] Kang Wangcheng, McAuley J. Self-attentive sequential recommendation[C]//Proceedings of 2018 IEEE International Conference on Data Mining. Singapore: IEEE, 2018: 197-206.

[23] Ma Jianxin, Zhou Chang, Cui Peng, et al. Learning disentangled representations for recommendation[C]//Proceedings of the 33rd International Conference on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2019: 513.

[24] 钱忠胜, 王亚惠, 俞情媛, 等. 利用伪重叠判定机制的多层循环GCN跨域推荐[J]. 软件学报, 2025, 36(9): 4327-4348.
Qian Zhongsheng, Wang Yahui, Yu Qingyuan, et al. Multi-layer recurrent GCN cross-domain recommendation with pseudo-overlap detection mechanism[J]. Journal of Software, 2025, 36(9): 4327-4348. (in Chinese)

[25] Zhou Haoyi, Zhang Shanghang, Peng Jieqi, et al. Informer: Beyond efficient transformer for long sequence time-series forecasting[C]//Proceedings of the 35th AAAI Conference on Artificial Intelligence. AAAI Press, 2021: 11106-11115.

[26] Zhu Yongchun, Chen Jingwu, Chen Ling, et al. AdaF^{2}M^2: Comprehensive learning and responsive leveraging features in recommendation system[C]//Proceedings of the 30th International Conference on Database Systems for Advanced Applications. Singapore: Springer, 2026: 332-343.

[27] Sang Lei, Wang Yu, Zhang Yiwen. Heterogeneous graph masked contrastive learning for robust recommendation[PP/OL]. V1. arXiv (2025-05-30)[2025-11-11]. https://arXiv.org/abs/2505.24172.

[28] Xia Xue, Eksombatchai P, Pancha N, et al. TransAct: Transformer-based realtime user action model for recommendation at pinterest[C]//Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. Long Beach: ACM, 2023: 5249-5259.

[29] Zhou Xiaofan, Lee K. ID and graph view contrastive learning with multi-view attention fusion for sequential recommendation[C]//Proceedings of 2024 IEEE International Conference on Big Data. Washington: IEEE, 2024: 690-699.

[30] Qiu Ruihong, Huang Zi, Yin Hongzhi, et al. Contrastive learning for representation degeneration problem in sequen-

tial recommendation[C]//Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining. New York: ACM, 2022: 813-823.
[31] Peng Furong, Liu Kang, Lu Xuan, et al. TSC: A simple two-sided constraint against over-smoothing[C]//Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. Barcelona: ACM, 2024: 2376-2387.
[32] Zhu Xiaotian, Zhou Wengang, Li Houqiang. Improving deep neural network sparsity through decorrelation regularization[C]//Proceedings of the 27th International Joint Conference on Artificial Intelligence. Stockholm: IJCAI, 2018: 3264-3270.
[33] Zeng Yuyuan, Dai Tao, Xia Shutao. Corrdrop: Correlation based dropout for convolutional neural networks[C]//Proceedings of 2020 IEEE International Conference on Acoustics, Speech and Signal Processing. Barcelona: IEEE, 2020: 3742-3746.
[34] Jin Wei, Liu Xiaorui, Ma Yao, et al. Feature overcorrelation in deep graph neural networks: A new perspective[C]//Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. Washington: ACM, 2022: 709-719.
[35] Lin Guanyu, Luo Jinwei, Li Yinfeng, et al. Iterative sparse attention for long-sequence recommendation[C]//Proceedings of the 39th AAAI Conference on Artificial Intelligence. Philadelphia: AAAI Press, 2025: 12147-12155.
[36] Fei Ke, Zhang Xinyue, Li Jingjing. Entire-space variational information exploitation for post-click conversion rate prediction[C]//Proceedings of the 39th AAAI Conference on Artificial Intelligence. Philadelphia: AAAI Press, 2025: 11654-11662.
[37] Wu Wei, Wang Chao, Shen Dazhong, et al. AFDGCF: Adaptive feature de-correlation graph collaborative filtering for recommendations[C]//Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval. Washington: ACM, 2024: 1242-1252.
[38] Devlin J, Chang Mingwei, Lee K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding[C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Minneapolis: ACL, 2019: 4171-4186.
[39] Tang Jiaxi, Wang Ke. Personalized top-N sequential recommendation via convolutional sequence embedding[C]//Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining. Marina Del Rey: ACM, 2018: 565-573.
[40] Sun Fei, Liu Jun, Wu Jian, et al. BERT4Rec: Sequential recommendation with bidirectional encoder representations from transformer[C]//Proceedings of the 28th ACM Iinternational Conference on Information and Knowledge Management. Beijing: ACM, 2019: 1441-1450.

作者简介

钱忠胜 男,1977年1月出生于江西省鹰潭市。2008年在上海大学获工学博士学位。江西财经大学教授、博士生导师。主要研究方向为软件工程、人工智能、推荐系统等。
E-mail: changesme@163.com

刘金平 男,1995年12月出生于江西省赣州市。江西财经大学计算机与人工智能学院博士研究生。主要研究方向为推荐系统、智能化软件工程等。
E-mail: 2202310083@stu.jxufe.edu.cn

李玉龙 男,1997年9月出生于山东省济宁市。江西财经大学计算机与人工智能学院博士研究生。主要研究方向为推荐系统、智能化软件工程等。
E-mail: 605786950@qq.com

范赋宇 男,2002年4月出生于江西省抚州市。江西财经大学计算机与人工智能学院硕士研究生。主要研究方向为推荐系统、智能化软件工程等。
E-mail: 2971549683@qq.com

陈 超 男,1999年9月出生于江西省九江市。江西财经大学计算机与人工智能学院硕士研究生。主要研究方向为推荐系统、智能化软件工程等。
E-mail: 2451284629@qq.com