研究与开发

基于时频变换与动态图注意力的多模态序列推荐方法

隋欣怡，王瑞琴，任宇彬，方驰
（湖州师范学院信息工程学院，浙江 湖州 313000）

摘　要：针对多模态序列推荐中用户兴趣演化建模、跨模态语义对齐及时频特征提取的不足，提出小波增强的动态图注意力推荐（wavelet-enhanced dynamic graph attention recommendation，Wave-DGARec）模型。该模型从时频变换、图建模与对比学习这3个维度进行创新设计：引入多尺度小波变换模块，对行为序列进行时频重构，捕捉非平稳偏好波动；构建用户–商品–图像三元动态图，利用图注意力机制实现结构化语义在不同模态间的高效传播；设计跨模态对比学习机制，引入可学习温度参数，提升语义对齐质量与样本判别能力。在Amazon 4个领域数据集的实验验证了Wave-DGARec的优越性。消融实验证实了小波模块与动态图建模的有效性，为多模态推荐系统提供了一种融合时频分析与结构建模的新范式。

关键词：多模态序列推荐；小波变换；动态图注意力；跨模态对比学习
中图分类号：TP391.3；TN919.3
文献标志码：A
doi: 10.11959/j.issn.1000−0801.DXKX250483

Multimodal sequential recommendation method based on time-frequency transformation and dynamic graph attention

Sui Xinyi, Wang Ruiqin, Ren Yubin, Fang Chi
School of Information Engineering, Huzhou University, Huzhou 313000, China

Abstract: To address limitations in modeling user interest dynamics, cross-modal semantic alignment, and time-frequency feature extraction in multimodal sequential recommendation, wavelet-enhanced dynamic graph attention recommendation (Wave-DGARec) was proposed. This framework introduced innovations in three dimensions. A multi-scale wavelet transformation module was introduced to reconstruct behavioral sequences in the time-frequency domain, enabling the capture of non-stationary preference fluctuations. A user-item-image tripartite dynamic graph was constructed, wherein a graph attention mechanism is leveraged to enable efficient propagation of structured semantics across different modalities. A cross-modal contrastive learning strategy with a learnable temperature parameter was designed, enhancing both semantic alignment and sample discrimination. Experimental results on four Amazon domain datasets demonstrated the superiority of Wave-DGARec. Ablation studies further validated the effectiveness of both

收稿日期：2025−07−30；**修回日期：**2025−12−30
通信作者：王瑞琴，wrq@zjhu.edu.cn
基金项目：国家自然科学基金资助项目（No.62277016）
Foundation Item: The National Natural Science Foundation of China (No.62277016)

the wavelet module and the dynamic graph modeling. This work introduced a novel paradigm for multimodal recommendation systems by seamlessly integrating time-frequency analysis with structured representation learning.

Key words: multimodal sequential recommendation, wavelet transform, dynamic graph attention, cross-modal contrastive learning

0 引言

随着图文信息在社交平台和电商应用中的持续增长，推荐系统[1]在电商场景中发挥着越来越重要的作用。基于项目ID的协同过滤方法[2]能挖掘用户历史行为中的潜在兴趣，但存在模态信息缺失与表达单一的问题。

多模态推荐系统通过融合文本、图像等模态信息[3]增强了语义表达，但在行为动态建模与模态间特征交互方面仍有不足。现有方法多采用模态拼接[4]或加权求和[5]的方式，难以刻画语境变化下用户兴趣的演化过程；频域建模依赖傅里叶变换[6]，但难处理局部非平稳的用户偏好；模态间语义关系常通过注意力机制或全连接层建模，缺乏结构化表达[7]。不同模态在表达用户意图时具有异质性，如何实现对语义的精准对齐及行为序列的时频建模，成为提升推荐性能的关键挑战。

为解决上述问题，本文提出小波增强的动态图注意力推荐（wavelet-enhanced dynamic graph attention recommendation，Wave-DGARec）模型，主要贡献如下。

（1）提出基于小波变换的时频变换模块，对文本和图像嵌入进行多尺度小波分解与重构，增强序列中用户行为的动态表达能力。

（2）构建用户–文本–图像三元异构图，引入图注意力机制，实现跨模态节点间的结构化语义传播。

（3）引入可学习温度参数的跨模态对比学习机制，优化正负样本之间的语义相似度分布，提升模型的表达能力和泛化性能。

（4）在多个公开数据集上的对比实验结果表明，所提方法在推荐准确性和鲁棒性方面均优于当前主流模型。

1 相关工作

1.1 序列推荐模型

序列推荐[8]通过分析用户的历史交互数据来预测用户未来的兴趣倾向或交互项目。基于门控循环单元的会话推荐（gated recurrent unit for session-based recommendation，GRU4Rec）模型利用循环神经网络（recurrent neural network，RNN）[9]捕捉用户行为中的时间依赖关系；Caser模型引入卷积神经网络（convolutional neural network，CNN）[10]建模局部时序模式；Transformer架构[11]进一步提升了序列建模的能力；SASRec[12]首次采用自注意力机制捕捉用户的长短期兴趣，但它仍依赖项目ID嵌入，缺乏模态语义的刻画。近年来，图神经网络逐渐应用于推荐任务，轻量级图卷积网络（light graph convolution network，LightGCN）[13]通过简化特征变换与激活操作，提升了邻接关系聚合的效率，代表了图卷积模型的轻量化发展趋势。

1.2 多模态推荐模型

为提升推荐系统的语义理解能力，研究者引入多模态信息来增强模型表征。视觉贝叶斯个性化排序（visual Bayesian personalized ranking，VBPR）[14]结合视觉特征与协同过滤，改善了冷启动表现；多模态图卷积网络（multi-modal graph convolutional network，MMGCN）[15]设计多路图神经网络，实现文本与图像的结构传播与融合；语义增强的语言模型推荐（semantic-

enhanced language model for recommendation，SLMRec）[3]利用预训练语言模型提升文本嵌入质量；面向推荐的多模态图卷积网络（multi-modal graph convolutional network for recommendation，MGCN）[16]统一建模多模态与行为节点，缓解模态冗余；FREEDOM[17]通过冻结部分模态图结构，减少冗余传播干扰。然而，这些方法在模态交互中普遍缺乏结构约束，限制了语义传播效率与模态互补的深度融合。对比学习[18]的引入进一步推动了多模态推荐的发展，通过构造正负样本对提升用户与项目表示的判别能力。例如，基于对比学习的跨模态推荐（contrastive learning-based cross-modal recommendation，CLCRec）[19]结合图结构与序列建模，引入对比目标缓解模态噪声与语义偏移，促进语义对齐。然而，在利用对比学习进行多模态推荐的相关方法中，多数方法仍采用固定温度参数，难以根据样本相似度动态调节，这在一定程度上影响了对比学习在多模态推荐中的性能提升效果。

2　模型设计

本文提出多模态序列推荐模型Wave-DGARec，该模型包含预处理模块、小波时频变换模块、动态图注意力模块、预测模块与跨模态对比学习模块。该模型首先对用户行为序列中的多模态信息进行预处理，生成统一低维表示；随后采用离散小波变换从频域角度获得行为序列的多尺度语义特征；引入图注意力机制刻画行为节点间的结构依赖与兴趣演化过程；引入跨模态对比学习目标，将融合后的表示与小波变换模态表示进行语义对齐；融合得到的多模态表示进入预测模块，输出最终推荐结果。基于时频变换与动态图注意力的多模态推荐架构如图1所示。

2.1　问题定义

在推荐系统中，用户集合记为U，项目集合记为I。每个项目$i \in I$均包含与之关联的图像信息与文本描述序列，文本描述由若干词元构成，记为$D_i=[w_1,w_2,\cdots,w_c]$，其中c表示文本序列的长度。每位用户$u \in U$拥有一个历史交互序列，定义为$S_u=[i_1,i_2,\cdots,i_T]$，其中$i_T \in I$表示用户在时间步t与项目的交互，序列中的项目按交互时间顺序排列，T为序列长度i_{T+1}。序列推荐任务的目标是基于用户的历史行为序列S_u，预测其在下一时刻可能交互的项目，该任务可形式化为学习一个排序函数，如下所示：

$$\hat{i}_{T+1}=\arg\max_{i \in I} f(S_u,i) \tag{1}$$

其中，$f(\cdot)$表示用户–项目对的偏好打分函数。

2.2　多模态预处理

为统一不同模态数据的维度，模型在结构建模前设计了预处理模块，分别对各模态进行编码

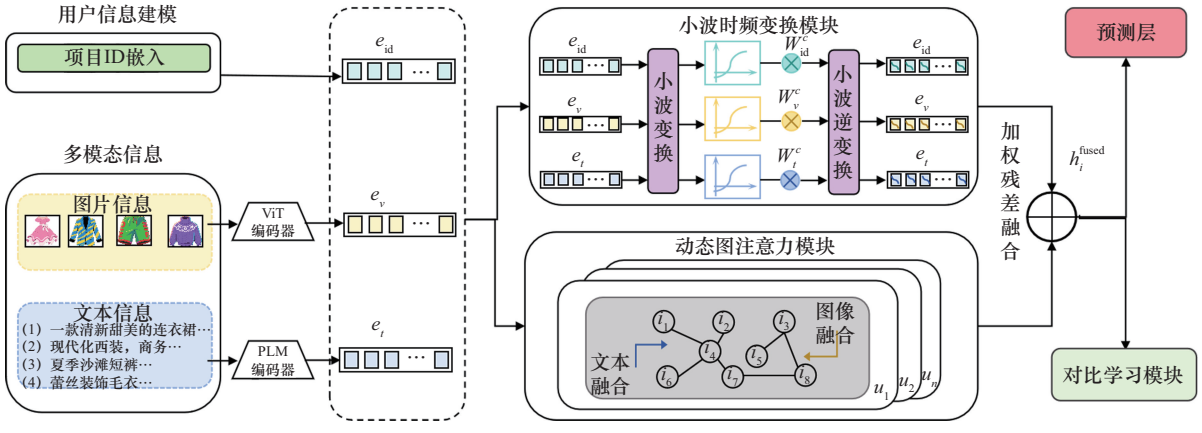


图1　基于时频变换与动态图注意力的多模态推荐架构

和格式规范。项目的文本信息通过预训练语言模型（pretrained language model，PLM）编码，在输入序列起始位置添加特殊标记[CLS]，聚合整段文本的语义信息。具体表示为：

$$t_j=\mathrm{PLM}\left([\mathrm{CLS}],\ w_1,\ w_2,\ w_3,\cdots,\ w_c\right) \tag{2}$$

为进一步增强模型对多样文本语义的建模能力，模型引入混合专家（mixture-of-experts，MoE）机制，对序列中每个位置的嵌入进行动态调制与特征转换，捕捉细粒度的语义差异。混合专家机制架构如图2所示。每个位置的融合嵌入e_t^j表示为：

$$e_t^j=\boldsymbol{G}\cdot\sum_{k=1}^{K}g_k\cdot\left(t_j+s_{j,k}\right)\cdot\boldsymbol{W}_P^k \tag{3}$$

其中，$\boldsymbol{G}$为全局门控矩阵，g_k为门控路由分配的权重，$s_{j,k}$为随机扰动项，$\boldsymbol{W}_P^k$为专家特征变换矩阵。最终构成文本模态的时序嵌入序列$E_t=\left[e_t^1,e_t^2,\cdots,e_t^T\right]$。

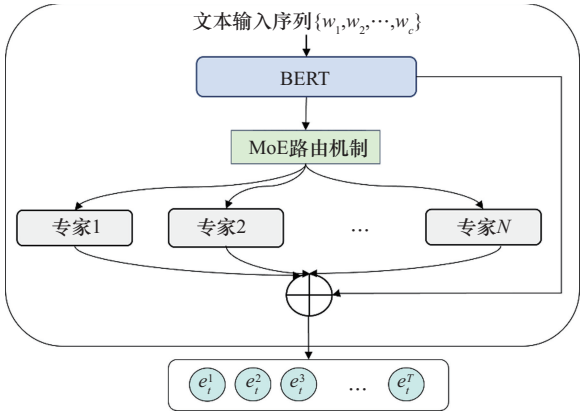


图2　混合专家机制架构

项目图像由图像编码器（vision transformer，ViT）处理为固定维度的向量形式。本文直接采用上游编码后的图像特征作为输入，记为$\boldsymbol{E}_v=\left[e_v^1,e_v^2,\cdots,e_v^T\right]\in\mathbf{R}^{T\times d}$。其中，$e_v^t\in\mathbf{R}^d$表示历史序列中第$t$个项目对应的图像表示。

除图像与文本模态外，模型还为每个项目分配了一个可学习的ID嵌入向量，具体表示为$\boldsymbol{e}_{\mathrm{id}}=\boldsymbol{P}[i]$，$\boldsymbol{P}\in\mathbf{R}^{|I|\times d}$，$\boldsymbol{P}$为项目ID嵌入矩阵，$i$为项目编号，$|I|$表示项目集合大小。

2.3　时频变换

为增强模型对模态序列中时间维度的频率结构语义建模能力，本文设计了时频变换模块，采用一维离散小波变换（discrete wavelet transform，DWT）进行通道级分解与逆变换建模。

为保证跨模态的结构一致性，各嵌入表示经过统一维度压缩，共享小波变换通道与配置参数。设输入模态序列为$\boldsymbol{E}^{B\times T\times d}$，其中$B$为batch size，$T$为序列长度，$d$为特征维度。为适配小波处理模块，模型首先进行维度压缩：

$$\boldsymbol{E}^{\mathrm{comp}}=\boldsymbol{E}\cdot\boldsymbol{W}^{\mathrm{comp}}+\boldsymbol{b}^{\mathrm{comp}},\ \boldsymbol{E}^{\mathrm{comp}}\in\mathbf{R}^{B\times T\times c} \tag{4}$$

其中，$\boldsymbol{E}^{\mathrm{comp}}\in\mathbf{R}^{B\times T\times c}$为压缩后的模态序列嵌入张量，作为小波处理模块的输入；$\boldsymbol{E}\in\mathbf{R}^{B\times T\times d}$为原始模态序列嵌入张量；$\boldsymbol{W}^{\mathrm{comp}}\in\mathbf{R}^{d\times c}$为小波处理维度，$c\ll d$。随后，对每个通道执行$L$级Daubechies小波分解，第$i$个通道的系数表示为：

$$\left\{C_i^{(l)}\right\}_{l=1}^{L}=\mathrm{DWT}^{(L)}\left(\boldsymbol{E}_i^{\mathrm{comp}}\right) \tag{5}$$

其中，L为小波分解层数，$C_i^{(l)}$表示第l层的频率子带信息。本文所选用的Daubechies中的db4小波函数具有良好的紧支撑性和边缘表达能力，在保持时域局部性的同时提取频率变化特征，适用于用户行为序列中非平稳偏好的建模。

完成频率建模后，模型将各通道的小波系数通过逆离散小波变换（inverse discrete wavelet transform，IDWT）还原为时域序列，通过线性扩展恢复原始嵌入维度，得到模态的频率增强表示：

$$\boldsymbol{H}^{\mathrm{wavelet}}=\mathrm{IDWT}\left(\left\{C_j^{(l)}\right\}\right)\cdot\boldsymbol{W}_{\mathrm{enlarge}}+\boldsymbol{b}_{\mathrm{enlarge}},\ \boldsymbol{H}^{\mathrm{wavelet}}\in\mathbf{R}^{B\times T\times d} \tag{6}$$

其中，$\boldsymbol{W}_{\mathrm{enlarge}}$为线性扩展权重矩阵，用于将小波处理后的低维表示恢复至原始嵌入维；$\boldsymbol{b}_{\mathrm{enlarge}}$为扩展过程中的偏置向量。最终输出的$\boldsymbol{H}^{\mathrm{wavelet}}$具备丰富的多尺度时频语义结构，保留了模态序列中隐含的行为波动与长期趋势特征，可在后续模块中与图结构建模结果进行融合。

2.4　动态图结构建模

为实现结构语义的时序演化建模，本文提出动态图结构建模模块，通过图注意力机制实现边权动态更新与语义传播。该模块以用户历史序列为基础，基于模态嵌入构建局部图结构，捕捉项目间的语义关联与多模态结构偏好。

图结构中的节点由项目 ID 嵌入 $e_{\mathrm{id}}^{t}\in\mathbf{R}^{d}$、文本模态 $e_{t}^{t}\in\mathbf{R}^{d}$、图像模态嵌入 $e_{v}^{t}\in\mathbf{R}^{d}$ 这 3 类表示构成。设用户行为序列长度为 T，每个项目在 3 个模态下的嵌入表示均经过线性映射统一为 d 维向量。3 类嵌入按模态维度拼接，形成图结构的节点集合：

$$V_u=\left[e_{\mathrm{id}}^{1},\cdots,e_{\mathrm{id}}^{T},e_{t}^{1},\cdots,e_{t}^{T},e_{v}^{1},\cdots,e_{v}^{T}\right],\quad V_u\in\mathbf{R}^{3T\times d} \tag{7}$$

为确保模态间信息充分传播，本文在局部序列范围内构建无向完全图，排除自环连接，边集合定义为：

$$E_u=\left\{(i,j)\mid i\neq j,\ i,j\in\left[1,\cdots,3T\right]\right\} \tag{8}$$

由于完全图的边数规模为 $O((3T)^2)$，在序列较长时可能带来计算开销，本文在设计中引入剪枝策略控制图结构规模，降低冗余传播带来的计算负担。剪枝依据包括模态相似度与历史交互强度，仅保留语义相关性较高的边连接，在保证信息覆盖的前提下提升计算效率。

图结构建模采用多头注意力机制实现跨模态交互与节点间语义传播。项目 ID 模态表示 $Q_u\in\mathbf{R}^{T\times d}$ 作为主查询路径，键值对输入分别为 $K_u,V_u\in\mathbf{R}^{3T\times d}$。结合剪枝后的边集合 E_u，输入图注意力模块后，仅对保留边所连接的节点对计算权重，实现邻居消息传递，计算式如下所示：

$$\boldsymbol{H}_{u}^{\mathrm{graph}}=\mathrm{GraphAttention}(Q_u,K_u,V_u,E_u) \tag{9}$$

虽然图拓扑在单次前向传播中保持为固定的完全图结构，但剪枝后的边集合与注意力权重会随模型参数动态变化。随着训练推进，模型通过参数驱动的边权更新与剪枝机制，自适应优化语义传播路径，虽未显式建模时间步间的递归演化，却有效刻画了用户兴趣的潜在变化趋势，实现了语义传播与结构表达的统一建模。

2.5　多路径残差融合

本文提出了多路径残差融合机制，将小波语义路径与动态图注意力模块输出进行加权整合，构建具备时频特征与结构感知能力的融合序列表示。同时，采用残差融合，缓解多路径信息引入的语义偏移，提高模型的表达完整性。

残差融合由小波频率语义路径 $\boldsymbol{H}_{u}^{\mathrm{wavelet}}$、项目 ID 与文本模态经过动态图注意力建模后的主结构路径 $\boldsymbol{H}_{u}^{\mathrm{graph}}$、图建模过程中的结构增强表示 $\boldsymbol{H}_{u}^{\mathrm{graph\text{-}img}}$ 这 3 类路径组成，将 3 路表示统一为相同维度后进行加权融合，融合表达定义为：

$$\begin{aligned}\boldsymbol{H}_{u}^{\mathrm{res}}=&(1+\omega_{\mathrm{wavelet}})\cdot\boldsymbol{H}_{u}^{\mathrm{wavelet}}+\\&\omega_{\mathrm{graph}}\cdot\boldsymbol{H}_{u}^{\mathrm{graph}}+\omega_{\mathrm{image}}\cdot\boldsymbol{H}_{u}^{\mathrm{graph\text{-}img}}\end{aligned} \tag{10}$$

融合后的序列表示 $\boldsymbol{H}_{u}^{\mathrm{res}}$ 作为最终嵌入特征，输入评分预测模块，用于候选项目的评分任务，有效提升了模型精度与泛化能力。考虑 3 条路径在建模目标与语义来源上的差异性，$\boldsymbol{H}_{u}^{\mathrm{wavelet}}$ 关注行为序列的频率动态，$\boldsymbol{H}_{u}^{\mathrm{graph}}$ 建模项目 ID 与文本模态的结构语义关系，$\boldsymbol{H}_{u}^{\mathrm{graph\text{-}img}}$ 引入图像模态在图结构中的补充信息，为实现灵活融合，本文引入权重系数 $\omega_{\mathrm{wavelet}}$、$\omega_{\mathrm{graph}}$、$\omega_{\mathrm{image}}$，分别控制 3 条路径在残差融合中的贡献程度。三者在语义空间中具备互补性，融合机制在提升表示多样性的同时，有效缓解了路径间的信息冗余与语义偏移问题。

2.6　跨模态对比学习

在完成多路径残差融合后，本文引入对比学习机制，对融合表示与小波频率语义路径之间的语义一致性进行监督优化，以缓解不同建模路径间的语义偏移问题。

残差融合后的序列表示 $\boldsymbol{H}_{u}^{\mathrm{res}}$ 作为查询向量，小波频率建模路径输出 $\boldsymbol{H}_{u}^{\mathrm{wavelet}}$ 作为正样本，构建基于语义路径的对比学习目标。从同一训练批次中通过随机采样采集其他用户的小波频率路径表示 $\boldsymbol{H}_{k}^{\mathrm{wavelet}}$ 构成负样本集合，定义路径对比损失为：

$$L_{\text{contrast}}=-\ln\frac{\exp\left(\frac{\text{sim}(\boldsymbol{H}_u^{\text{res}},\boldsymbol{H}_u^{\text{wavelet}})}{\tau}\right)}{\exp\left(\frac{\text{sim}(\boldsymbol{H}_u^{\text{res}},\boldsymbol{H}_u^{\text{wavelet}})}{\tau}\right)+\sum_{k\neq u}\exp\left(\frac{\text{sim}(\boldsymbol{H}_u^{\text{res}},\boldsymbol{H}_k^{\text{wavelet}})}{\tau}\right)} \tag{11}$$

其中，$\text{sim}(\cdot,\cdot)$为余弦相似度函数，$\tau\in\mathbf{R}^+$为可学习温度参数，用于调节对比目标的分布密度，正负样本分别来自同一用户与其他用户的小波频率路径输出。

2.7 模型训练

本文在训练阶段采用温度调控的交叉熵损失函数，融合路径级对比学习目标进行联合优化。训练目标包含两部分：主损失用于监督候选项目的排序结果；对比损失用于增强频率语义路径在融合表示中的主导性。

主损失基于候选项目的归一化评分，对用户与候选项目之间的匹配程度进行建模。构建信息增益型排序监督信号，具体定义为：

$$L_{\text{main}}=-\ln\left(\frac{\exp\left(\frac{s(u,\tau^+)}{\tau}\right)}{\sum_{j\in C}\exp\left(\frac{s(u,j)}{\tau}\right)}\right) \tag{12}$$

其中，$s(u,j)$表示用户u对候选项目j的预测得分，由融合后的序列表示与项目多模态表示计算得到，C为用户对应的候选项目集合，$\tau\in\mathbf{R}^+$为温度系数，用于调节softmax分布的平滑程度。

同时，为强化融合表示与频率语义主干之间的对齐一致性，本文引入路径级对比学习损失L_{contrast}，其定义见式(11)，最终训练目标为：

$$L_{\text{total}}=L_{\text{main}}+\lambda\cdot L_{\text{contrast}} \tag{13}$$

其中，λ为可调超参数。该联合训练机制引导模型在保持语义结构一致性的同时，增强了模态融合后的表征判别性与鲁棒性。

3 实验结果

3.1 数据集

本文采用Amazon推荐数据集中的4个典型子集进行实验评估，分别为Baby、Sports、Clothing和Electronics。该数据集广泛用于多模态推荐研究，适用于验证融合建模与时序结构建模的泛化能力。为提升序列建模的有效性与模态融合的稳定性，实验阶段采用如下预处理策略：对用户与项目执行5-core筛选，移除缺失图像或文本的项目，并使用Leave-One-Out采样构造数据集。数据集统计信息见表1。

表1 数据集统计信息

数据集	用户数量	项目数量	交互次数	稀疏性
Baby	19 445	7 050	160 792	99.88%
Sports	35 598	18 357	296 337	99.95%
Clothing	39 387	23 033	278 677	99.97%
Electronics	192 403	63 001	1 689 188	99.99%

3.2 评价指标

为全面评估模型在序列推荐任务中的排序性能，本文选取2类主流指标进行结果对比与性能分析，分别为Recall@K和NDCG@K，其中$K\in\{10,20\}$。

3.3 基准模型

为评估所提方法的性能，本文选取以下7个具有代表性的推荐模型作为对比基线。

（1）BPR[14]：利用用户–项目交互数据进行隐式偏好建模。

（2）LightGCN[13]：提出简化的图卷积结构，通过邻接传播构建用户与项目表示。

（3）VBPR[14]：在BPR基础上引入图像视觉模态信息，增强用户偏好学习。

（4）MMGCN[15]：基于图神经网络框架，融合文本与图像等多模态信息。

（5）SLMRec[3]：采用Transformer架构建模用户行为序列，融合多模态特征。

（6）MGCN[16]：在图神经网络中注入行为序列信号，实现用户兴趣与多模态特征的联合学习。

（7）FREEDOM[17]：采用双重优化机制，优化主任务目标与模态协同机制。

3.4　对比结果

不同模型之间的性能比较结果见表 2，最佳性能结果加粗显示，次优性能采用下划线标记。为验证性能差异的统计显著性，本文在 5 组随机种子下对 Recall@K 与 NDCG@K 进行配对 t 检验，结果显示本文模型在多个指标上显著优于最优基线（p<0.01），体现出稳定性与统计可信度。综合各项指标表现，本文模型在多个数据集上展现出较好的推荐效果：相较 BPR 和 LightGCN，本文模型能够有效捕捉用户潜在偏好，在稀疏交互场景下保持较高的准确率；与 VBPR、MMGCN 等融合模态模型相比，本文模型在模态表示学习上具有更强能力，尤其体现在 Clothing 与 Electronics 数据集上。值得注意的是，在 Sports 数据集上，Recall@20 指标略逊于现有最优模型，这可能是因为用户兴趣分布广泛、行为模式差异较大，提升了建模难度。

3.5　消融实验

当前多模态推荐方法多聚焦于模态融合或单一结构建模，在频率建模与语义一致性约束方面仍存在不足，难以有效捕捉用户行为中的非平稳特征与跨模态语义偏移。为了验证 WaveDGARec 中频率建模、结构建模与语义对齐机制的协同作用，本文设计了以下 4 个变体开展消融实验。

（1）w/o Wavelet：移除小波变换模块，使用原始模态表示参与后续图建模。

（2）w/o Graph Attention：替换动态图注意机制为静态图卷积操作。

（3）w/o Contrast：取消对比学习损失，使用单一分类目标进行优化。

（4）Simple Fusion：采用加权平均策略整合模态表示。

为了确保公平比较，本文统一采用 Adam 优

表 2　不同模型之间的性能比较结果

数据集	评价指标	通用推荐模型		多模态特征的序列模型					本文模型
		BPR	LightGCN	VBPR	MMGCN	SLMRec	MGCN	FREEDOM	
Baby	Recall@10	0.038 2	0.045 3	0.042 5	0.042 4	0.054 5	0.061 6	0.062 4	**0.064 6**
	Recall@20	0.059 5	0.07 28	0.066 3	0.066 8	0.083 7	0.093 4	0.098 5	**0.099 8**
	NDCG@10	0.020 7	0.024 6	0.022 3	0.022 3	0.029 6	0.033 0	0.032 4	**0.035 1**
	NDCG@20	0.026 3	0.031 7	0.028 4	0.028 6	0.037 1	0.041 4	0.041 6	**0.042 7**
Sports	Recall@10	0.041 7	0.054 2	0.056 1	0.038 6	0.067 6	0.072 6	0.072 4	**0.073 4**
	Recall@20	0.063 3	0.083 7	0.085 7	0.062 7	0.101 7	**0.110 5**	0.108 9	0.110 4
	NDCG@10	0.023 2	0.030 0	0.030 7	0.020 4	0.037 4	0.040 1	0.039 0	**0.040 7**
	NDCG@20	0.028 8	0.037 6	0.038 4	0.026 6	0.046 2	0.049 4	0.048 4	**0.050 2**
Clothing	Recall@10	0.020 0	0.033 8	0.028 1	0.022 4	0.046 1	0.064 9	0.063 5	**0.066 1**
	Recall@20	0.029 5	0.051 7	0.041 0	0.036 2	0.069 6	0.097 1	0.093 8	**0.098 4**
	NDCG@10	0.011 1	0.018 5	0.015 7	0.011 8	0.024 9	0.040 3	0.034 0	**0.042 5**
	NDCG@20	0.013 5	0.023 0	0.019 0	0.015 3	0.030 8	0.049 8	0.041 7	**0.051 2**
Electronics	Recall@10	0.023 5	0.036 3	0.029 3	0.020 7	0.036 7	0.043 9	0.038 2	**0.045 7**
	Recall@20	0.036 7	0.054 0	0.045 8	0.033 1	0.057 3	0.064 3	0.058 8	**0.065 5**
	NDCG@10	0.012 7	0.020 4	0.015 9	0.010 9	0.019 8	0.024 5	0.020 9	**0.025 8**
	NDCG@20	0.016 1	0.025 0	0.020 2	0.014 1	0.025 9	0.029 8	0.026 2	**0.031 0**

化器和交叉熵损失函数来优化模型，训练批次大小设为2 048。若验证集上的NDCG@10在10个epoch内无提升，则采用提前停止策略防止过拟合。消融实验结果如图3所示。图3展示了5种配置在Recall@10与NDCG@10上的对比结果。结果显示，Wavelet模块提升了模型对行为波动的响应能力，Graph Attention增强了结构感知与语义传播，对比学习机制有效缓解了多路径融合中的语义偏移，Simple Fusion性能显著低于完整模型，反映出各模块在整体架构中的协同作用。

从机制层面看，Wavelet提供频域压缩与局部增强，改善模态输入；Graph Attention建模结构上下文依赖，实现语义传播；对比学习通过最大化正负样本间互信息，强化语义一致性。三者构成互补路径：Wavelet提供区分性输入，助力结构建模；Graph Attention构建语义边界，支撑对比学习；对比学习反向优化模态表示的一致性，提升频域建模稳定性。三者依次优化表示增强、结构建模与语义对齐，形成闭环机制。消融实验结果表明，任一模块移除均导致性能下降，验证了该结构对多模态推荐关键维度的全面覆盖与优越性能。

3.6 效率分析

为评估模型在不同数据规模下的计算开销，本文选取VBPR、MMGCN、MGCN与FREEDOM这4种具有代表性的推荐方法作为对比基线。在4个数据集上统计训练时间与参数量，结合Recall@10指标进行综合评估。不同模型的效率对比见表3。由表3可知，在训练时间方面，Wave-DGARec在Baby数据集上的训练效率优于MGCN，低于MMGCN和FREEDOM，但整体表现仍处于可接受范围，展现出一定的可扩展性。在参数量方面，Wave-DGARec的模型规模保持在中等水平，体现出结构设计上的紧凑性。相较于部分计算开销更低的模型，Wave-DGARec在Recall@10上表现更优，这说明其在不增加复杂度的前提下实现了性能提升，验证了其在多数据规模与复杂交互场景下的可行性与高效性。

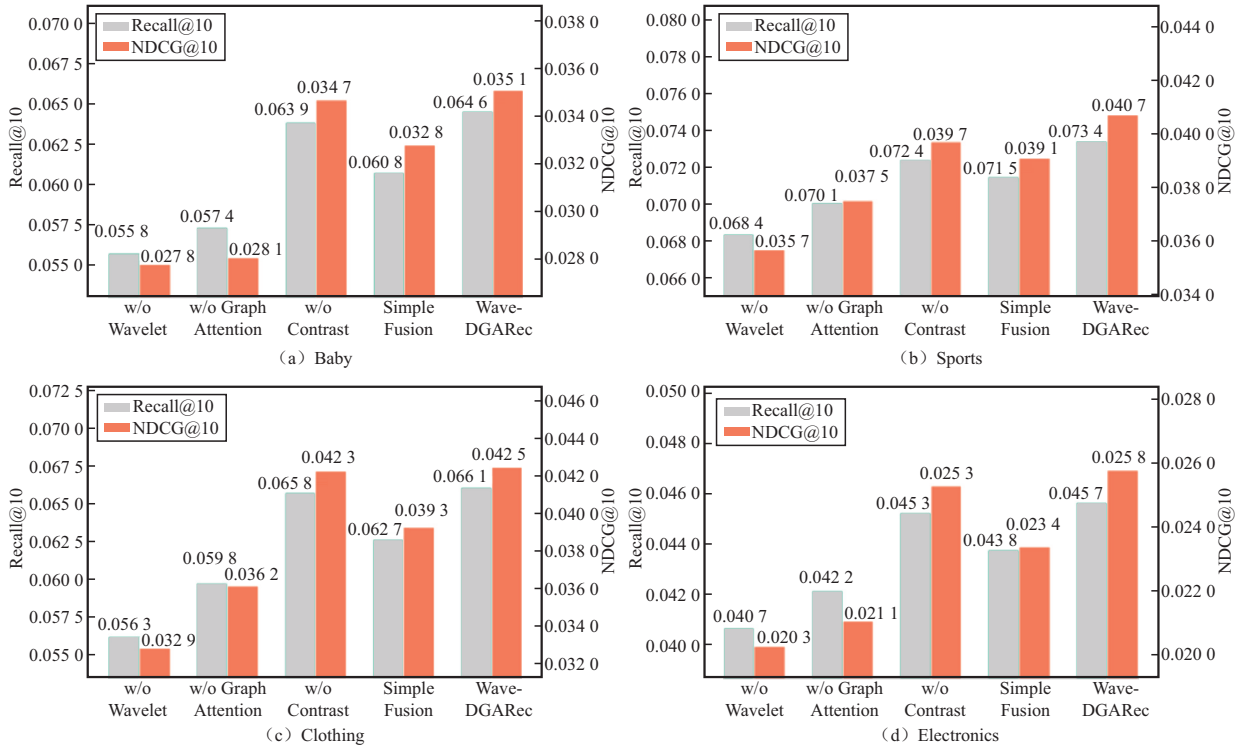


图3 消融实验结果

表3　不同模型的效率对比

数据集	方法	VBPR	MMGCN	MGCN	FREEDOM	Wave-DGARec
Baby	训练时间/(s·epoch^{-1})	1.12	5.68	10.25	2.77	8.02
	参数量/M	3.20	6.91	1.98	8.39	4.16
	Recall@10	0.042 5	0.042 4	0.061 6	0.062 4	**0.064 6**
Sports	训练时间/(s·epoch^{-1})	2.55	16.65	22.16	5.86	14.78
	参数量/M	6.00	12.84	3.79	16.62	8.77
	Recall@10	0.056 1	0.038 6	0.072 6	0.072 4	**0.073 4**
Clothing	训练时间/(s·epoch^{-1})	2.9	17.32	24.50	6.29	15.03
	参数量/M	6.8	13.32	4.30	18.03	9.69
	Recall@10	0.028 1	0.022 4	0.064 9	0.063 5	**0.066 1**
Electronics	训练时间/(s·epoch^{-1})	14.2	106.5	104.7	31.9	90.6
	参数量/M	6.40	65.65	16.65	98.40	58.80
	Recall@10	0.029 3	0.020 7	0.043 9	0.038 2	**0.045 7**

3.7　可视化分析

为验证小波模块对非平稳行为的建模能力，本文绘制频率–时间热力图进行可视化分析。实验采用 Daubechies 4（db4）小波对压缩特征序列进行三级分解，提取低频（Low）、中频（Mid）、高频（High）系数，并结合注意力机制构建响应图，横轴为时间步，纵轴为频率段，色条范围设为[0,0.5]。频率–时间二维热力图如图 4 所示。由图 4 可知，Baby 数据集在第 10～12 步高频段注意力增强，表明模型能识别集中型用户的突发行为；Sports 数据集中频段呈周期性抬升，体现对重复行为的分段关注；Clothing 数据集高频段在

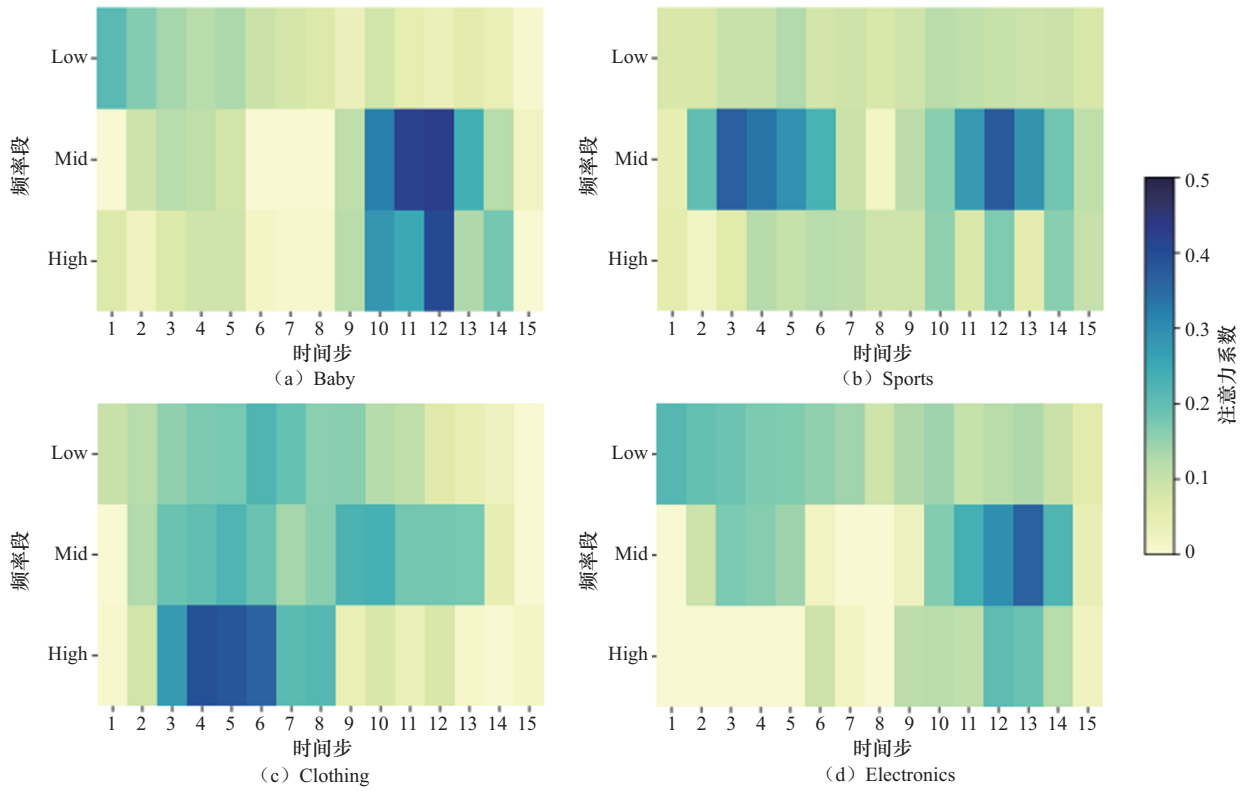


图4　频率-时间二维热力图

第4～6步轻微聚焦，中频响应平稳，说明模型能捕捉细腻行为；Electronics数据集高频响应较弱，中频在第12～14步略有提升，符合稀疏型用户的稳定特征。

模型在不同频率段上形成了差异化的注意力分布，具备一定的非平稳建模能力，尤其在高频帧处的聚焦表现出一定的敏感性，但部分频段响应仍存在平滑过度或时间步聚焦偏集中的现象，后续会考虑多尺度注意力等策略，以提升模型对突发行为的识别精度。

3.8 参数敏感性分析

为探究关键参数对模型性能的影响，本文对专家数量、小波压缩层数、温度系数及对比学习的λ系数进行了敏感性实验。评估指标为Recall@10和NDCG@10。

首先，本文设置MoE数量为2、4、6、8、10，MoE数量对模型性能的影响如图5所示。由图5可知，随着MoE数量的增加，Recall@10和NDCG@10整体呈上升趋势，说明专家机制有助于增强模型的表达能力。当专家数量超过8时，性能略有下降，表明过多专家可能引入冗余或干扰特征聚合，影响模型的泛化。

为评估小波压缩层数的影响，本文设置压缩层数为1、2、4、6、8，压缩层数对模型性能的影响如图6所示。由图6可知，随着层数加深，模型性能逐步下降，Clothing与Sports数据集尤为明显，表明过深压缩可能导致语义信息损耗，削弱频域建模效果。相比之下，Baby数据集表现较平稳，暗示其用户兴趣更集中，对语义粒度要求较低。

为评估温度系数τ的影响，本文设置温度参数为0.01、0.03、0.05、0.07、0.10。温度系数τ对模型性能的影响如图7所示。由图7可知，在Baby数据集上，Recall@10在τ=0.05和τ=0.07达到近似峰值时，性能差异较小，体现模型在中等温度区间的稳定性。当τ=0.01、0.03和0.1时性能

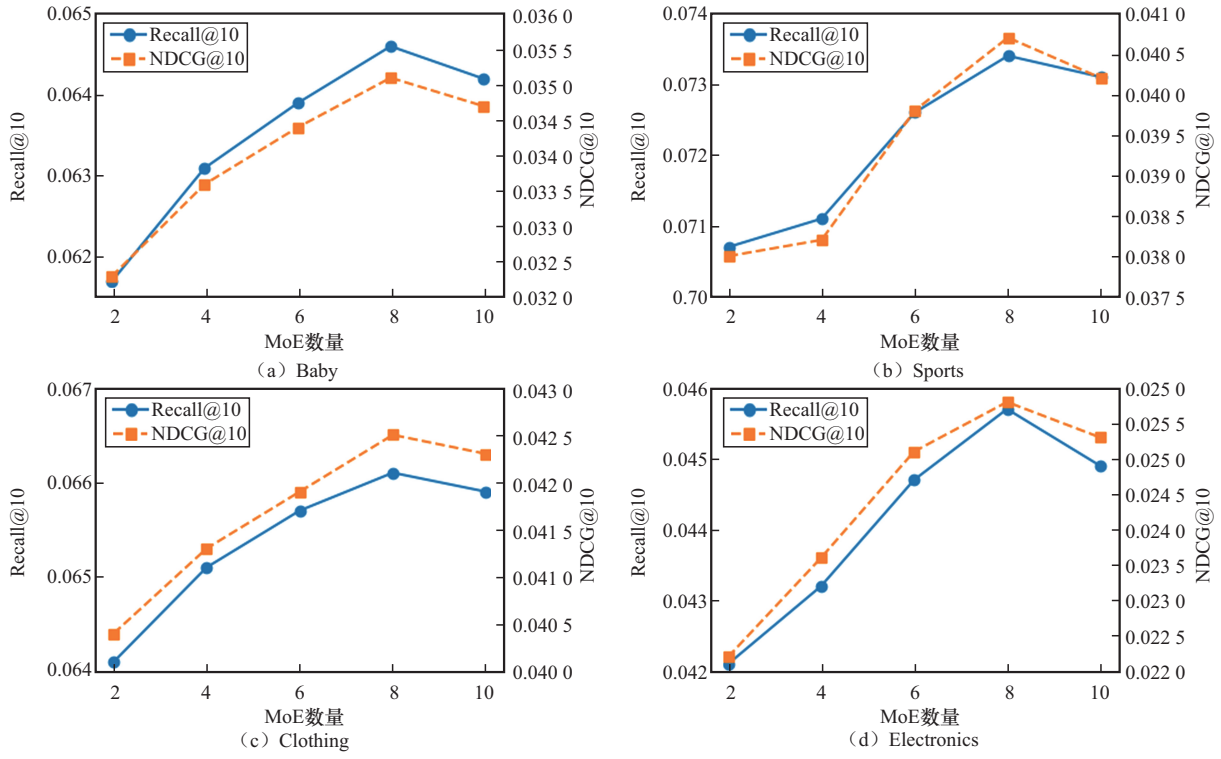


图5 MoE数量对模型性能的影响

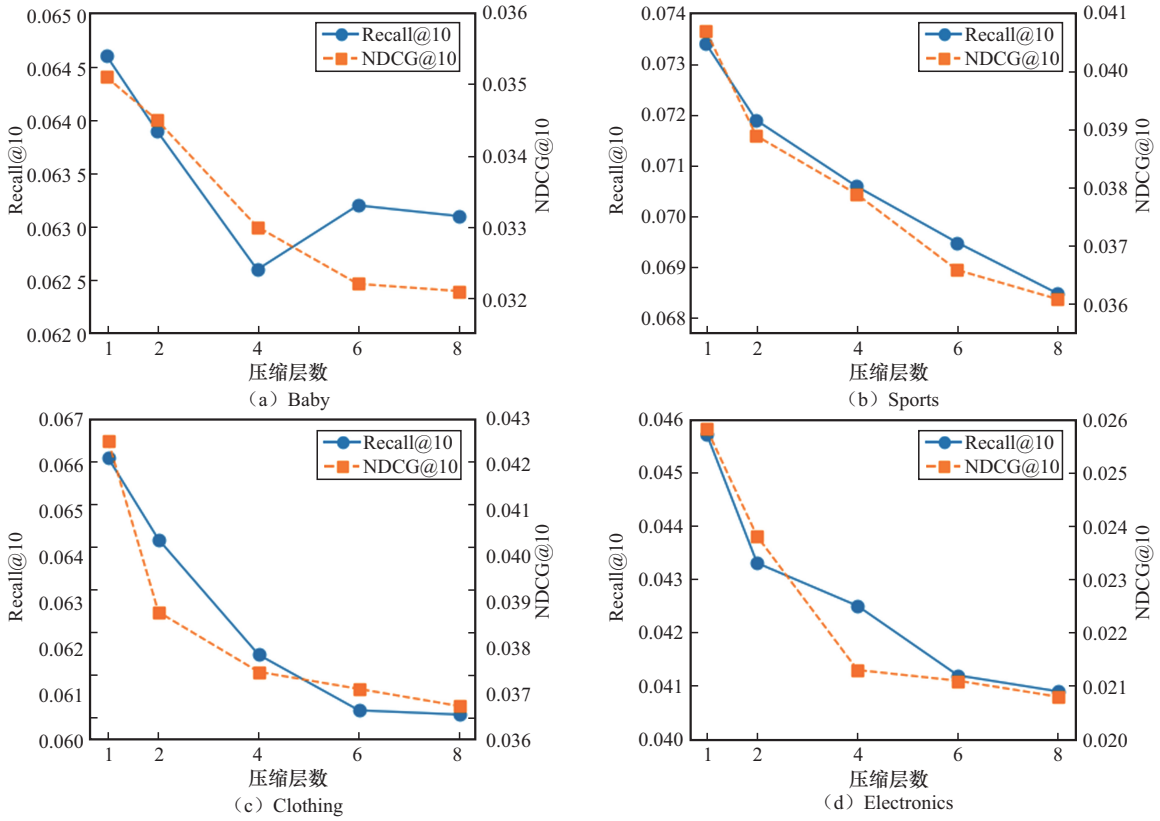


图6　压缩层数对模型性能的影响

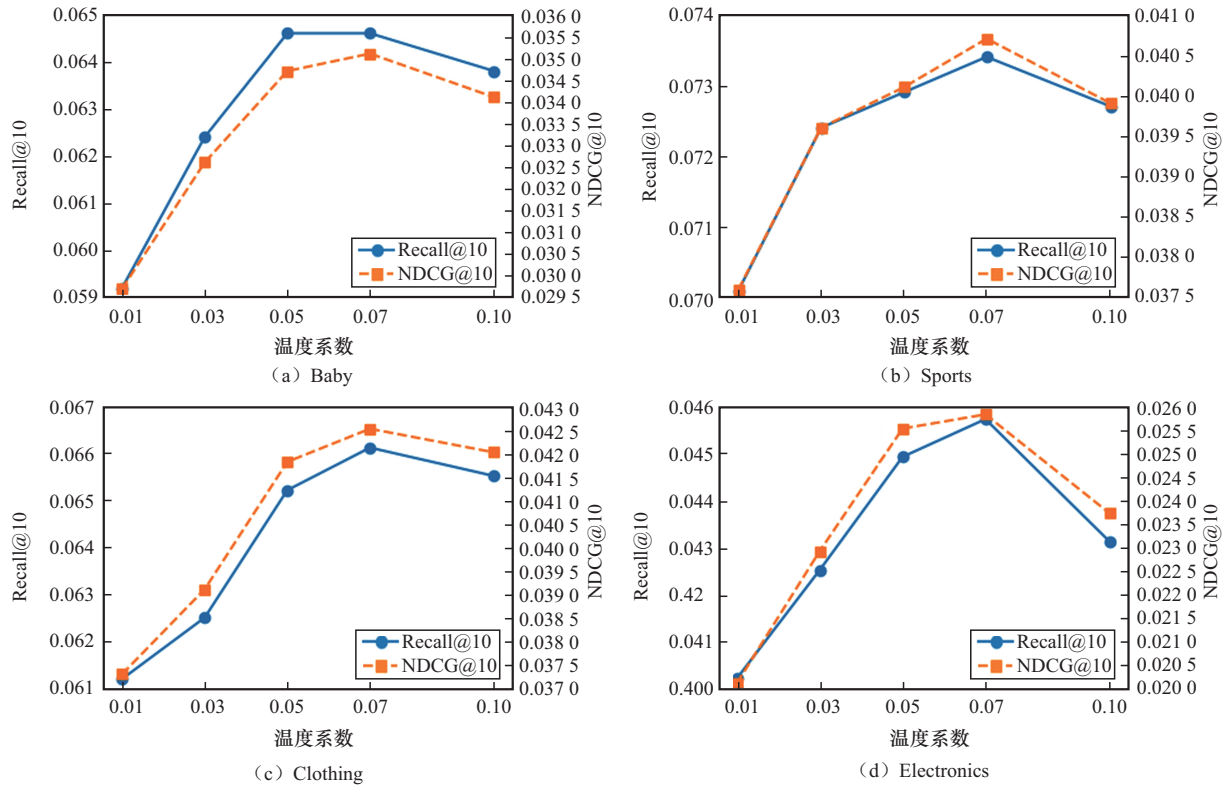


图7　温度系数τ对模型性能的影响

明显下降，验证了过低温度导致表示过度集中、过高温度引入噪声的假设。综合考虑各项指标，τ=0.07 在 Baby 数据集上为最优选择，但 τ=0.05 也具备稳定且接近最优的性能，体现了模型在中等温度区间的鲁棒性。整体来看，Recall@10 和 NDCG@10 在温度系数为 0.07 时达到峰值，表明适度的温度调控可增强序列表示的区分性。系数过低导致表示过度集中，系数过高引入噪声影响优化效果。

最后，本文研究了对比学习系数 λ 对推荐性能的影响，设置 λ 为 0.001、0.003、0.01、0.03、0.1。对比学习系数 λ 对模型性能的影响如图 8 所示。由图 8 可知，随着 λ 的增加，模型性能先升后降。在 Baby 数据集上，NDCG@10 在 λ=0.03 时略有回升，可能是因为对比目标调整了部分样本的表示结构，提升了排序得分，但整体优化仍不稳定。当 λ=0.01 时性能最佳，超过该值后对比目标可能主导训练，削弱主任务信号，导致推荐性能下降。

4 结束语

本文针对多模态序列推荐方法在特征表达与用户兴趣建模方面的不足，提出了 Wave-DGARec 模型。该模型通过多尺度小波变换刻画用户行为的时频动态特征，结合动态图神经网络实现跨模态语义传播与融合，并引入路径级对比学习与可学习温度参数，提升模态间的语义一致性与判别能力。

实验结果表明，Wave-DGARec 在 4 个公开数据集上均优于现有方法，验证了其在多模态序列推荐中的有效性与推广潜力。未来可进一步探索其在兴趣迁移建模及大语言模型融合方面的应用。

参考文献：

[1] Kumar C, Chowdary C R, Meena A K.Recent trends in recommender systems: a survey[J].International Journal of Multimedia Information Retrieval, 2024, 13(4): 41.

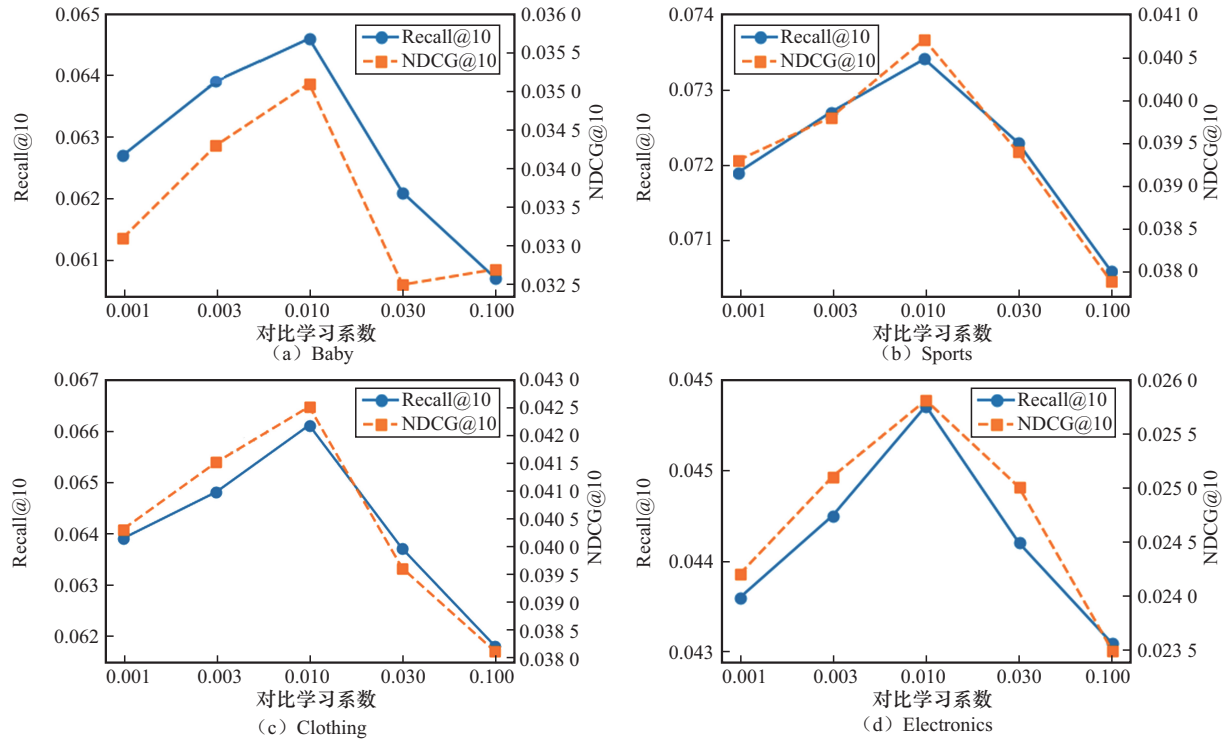


图 8 对比学习系数 λ 对模型性能的影响

[2] Chen X, Xu H T, Zhang Y F, et al. Sequential recommendation with user memory networks[C]//Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining. New York: ACM Press, 2018: 108-116.

[3] Tao Z L, Liu X H, Xia Y W, et al. Self-supervised learning for multimedia recommendation[J]. IEEE Transactions on Multimedia, 2023, 25: 5107-5116.

[4] Zheng Z, Chao W S, Qiu Z P, et al. Harnessing large language models for text-rich sequential recommendation[C]//Proceedings of the ACM Web Conference 2024. New York: ACM Press, 2024: 3207-3216.

[5] Liu Z H, Mei S, Xiong C Y, et al. Text matching improves sequential recommendation by reducing popularity biases[C]//Proceedings of the 32nd ACM International Conference on Information and Knowledge Management. New York: ACM Press, 2023: 1534-1544.

[6] Ong R K, Khong A W H. Spectrum-based modality representation fusion graph convolutional network for multimodal recommendation[C]//Proceedings of the Eighteenth ACM International Conference on Web Search and Data Mining. New York: ACM Press, 2025: 773-781.

[7] Deng H Y, Hu J L. Improving sequential recommendation with global item transitions and local subsequences[J]. IEEJ Transactions on Electrical and Electronic Engineering, 2024, 19(1): 100-108.

[8] 李家乐, 王瑞琴, 于洋. 数据增强的多模式时间感知序列推荐[J]. 电信科学, 2024, 40(11): 66-78.
Li J L, Wang R Q, Yu Y. Multi-pattern time-aware sequential recommendation with data augmentation[J]. Telecommunications Science, 2024, 40(11): 66-78.

[9] Pankratz M, Lee H, LAD L, et al. GRU4RecBE: a hybrid session-based movie recommendation system (student abstract) [C]//Proceedings of the Thirty-Sixth AAAI Conference on Artificial Intelligence, Thirty-Fourth Conference on Innovative Applications of Artificial Intelligence, and Twelfth Symposium on Educational Advances in Artificial Intelligence. New York: ACM Press, 2022: 13029-13030.

[10] Tang J X, Wang K. Personalized top-N sequential recommendation via convolutional sequence embedding[C]//Proceedings of the 11th ACM International Conference on Web Search and Data Mining. New York: ACM Press, 2018: 565-573.

[11] 王瑞琴, 黄熠旻, 纪其顺, 等. 注意力感知的边–节点交换图神经网络模型[J]. 电信科学, 2024, 40(1): 106-114.
Wang R Q, Huang Y M, Ji Q S, et al. Attention aware edge-node exchange graph neural network[J]. Telecommunications Science, 2024, 40(1): 106-114.

[12] Kang W C, Mcauley J. Self-attentive sequential recommendation[C]//Proceedings of the 2018 IEEE International Conference on Data Mining (ICDM). Piscataway: IEEE Press, 2018: 197-206.

[13] He X N, 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. New York: ACM Press, 2020: 639-648.

[14] He R N, Mcauley J. VBPR: visual Bayesian personalized ranking from implicit feedback[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2016, 30(1): 144-150.

[15] Wei Y W, Wang X, Nie L Q, et al. MMGCN: multi-modal graph convolution network for personalized recommendation of micro-video[C]//Proceedings of the 27th ACM International Conference on Multimedia. New York: ACM Press, 2019: 1437-1445.

[16] Geng X, Li Y G, Wang L Y, et al. Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting[C]//Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence. New York: ACM Press, 2019: 3656-3663.

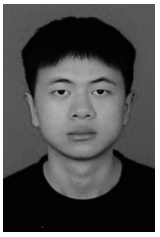
[17] Zhou X, Shen Z Q. A tale of two graphs: freezing and denoising graph structures for multimodal recommendation[C]//Proceedings of the 31st ACM International Conference on Multimedia. New York: ACM Press, 2023: 935-943.

[18] 于洋, 王瑞琴. 基于对比增强时间感知自注意力机制的序列推荐[J]. 电信科学, 2025, 41(1): 137-147.
Yu Y, Wang R Q. Sequential recommendation based on contrast enhanced time-aware self-attention mechanism[J]. Telecommunications Science, 2025, 41(1): 137-147.

[19] Xie X, Sun F, Liu Z Y, et al. Contrastive learning for sequential recommendation[C]//Proceedings of the 2022 IEEE 38th International Conference on Data Engineering (ICDE). Piscataway: IEEE Press, 2022: 1259-1273.

[作者简介]

隋欣怡（2000– ），女，湖州师范学院信息工程学院硕士生，主要研究方向为深度学习、自然语言处理、个性化推荐。



任宇彬（2000– ），男，湖州师范学院信息工程学院硕士生，主要研究方向为个性化推荐、数据挖掘。

王瑞琴（1979– ），女，博士，湖州师范学院信息工程学院教授，主要研究方向为自然语言处理、社交网络分析、个性化推荐。

方驰（2000– ），男，湖州师范学院信息工程学院硕士生，主要研究方向为序列推荐、多模态推荐。