

大语言模型增强的时间注意力推荐系统

孙浩然¹, 王欣², 熊菲¹

(1. 北京交通大学电子信息工程学院, 100044, 北京; 2. 北京计算机技术及应用研究所, 100039, 北京)

摘要: 为了解决传统推荐方法依赖于用户与物品的稀疏交互数据,其难以深入挖掘用户偏好背后的语义逻辑及其随时间演化的动态特征,以及大语言模型(LLM)在推荐系统中的直接应用受限于缺乏结构化交互建模与时间敏感性考虑的问题,提出了一种大语言模型增强的基于时间的推荐模型(LLATR)。LLATR旨在融合语义理解能力与用户兴趣的时间变化建模,以提升推荐精度与系统响应的个性化水平。模型设计了协同特征提取网络和时间特征建模网络,以及结合大语言模型生成的语义评分向量,通过对比学习机制,实现了多模态信息的统一建模,从而构建了具有动态适应能力的推荐框架。结果表明:在 MovieLens-100 K、Kaggle-Movie 这 2 个数据集上,LLATR 的均方根误差、平均绝对误差相对于现有主流模型提升了2%~5%。进一步分析表明,LLM能够补充协同特征之外的深层语义信息,并增强推荐系统对冷启动用户、稀疏数据和复杂行为背景的适应能力,以及有效建模用户的兴趣随时间的非线性演化趋势。

关键词: 大语言模型;推荐系统;评分预测;时间注意力推荐

中图分类号: TP391 **文献标志码:** A

DOI: 10.7652/xjtuxb202510021 **文章编号:** 0253-987X(2025)10-0221-10

Large Language Model-Augmented Time-Attention Recommender Systems

SUN Haoran¹, WANG Xin², XIONG Fei¹

(1. School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing 100044, China;

2. Beijing Institute of Computer technology and Application, Beijing 10039, China)

Abstract: To address the limitations of traditional recommendation methods that rely on sparse user-item interaction data—which struggle to deeply mine the semantic logic behind user preferences and their dynamic temporal evolution—as well as the constraints of directly applying large language models (LLMs) in recommendation systems due to their lack of structured interaction modeling and temporal sensitivity, a large language model-augmented time-aware recommendation (LLATR) algorithm is proposed. This algorithm aims to integrate semantic comprehension capabilities with temporal modeling of user interests to enhance recommendation accuracy and system personalization. The method designs a collaborative feature extraction network and a time feature modeling network, and combines the semantic scoring vectors generated by LLMs. Through a contrastive learning mechanism, it achieves unified modeling of multimodal information, thereby constructing a dynamically adaptive recommendation framework. Experiments are conducted on two datasets: MovieLens-100 K and Kaggle-Movie. The results demonstrate that LLATR improves root mean square error (RMSE) and mean absolute error (MAE) by 2%—5%

compared to existing mainstream models. Further analysis reveals that LLMs can supplement deep semantic information beyond collaborative features, enhance the recommendation system's adaptability to cold-start users, sparse data, and complex behavioral contexts, and effectively model the nonlinear evolution trends of user interests over time.

Keywords: large language modeling; recommender systems; rating prediction; time-attention recommender

随着信息技术的快速发展,互联网数据量呈现爆炸式增长,用户面临信息过载的问题日益严重。为解决这一问题,推荐系统被广泛应用于电子商务、社交媒体和内容平台中,以帮助用户从大量数据中筛选出个性化和相关性较强的信息。然而,随着用户需求的动态变化和数据规模的不断扩大,如何提高推荐系统的个性化和时效性,尤其是如何有效利用时间信息,成为当前推荐系统研究中的重要课题。

传统推荐系统主要包括基于内容的推荐、协同过滤、序列推荐和混合推荐等方法。基于内容的方法通过提取物品和用户特征实现匹配^[1-2],协同过滤基于用户-物品交互历史进行推荐^[3-4],而序列推荐进一步关注用户行为的时间序列,捕捉兴趣的动态演变^[5]。然而,这些方法往往忽视了用户与物品交互时间的重要性。现有的时间序列模型虽然能够建模用户行为序列,但未能有效区分不同时间段交互对用户兴趣变化的影响,导致在兴趣快速变化的场景中表现不佳^[6]。

近年来,大语言模型(LLM)在自然语言处理领域取得突破,并在语义理解、推理和生成任务上展现出强大能力^[7]。尽管如此,LLM在推荐系统中的应用仍面临挑战,其主要优势在于文本信息的语义理解^[8],而传统推荐任务则多基于用户-物品ID交互数据,LLM难以直接建模这些离散ID间的关系^[9]。此外,由于LLM缺乏对协同过滤信息的有效建模,如何将其强大的语义理解能力与推荐系统中的用户行为建模相结合,仍是一个亟待解决的问题^[10]。

针对上述挑战,本文提出了一种基于时间增强的大语言模型推荐框架(LLATR),通过引入用户与物品交互的时间信息,优化LLM在推荐任务中的应用。LLATR结合时间序列建模与语义特征提取,不仅能够捕捉用户兴趣的动态变化,还能有效提升推荐的个性化和时效性。在真实数据集上的实验验证了LLATR框架的有效性,相较于现有方法,其推荐准确性和用户满意度均取得了显著提升。

1 大语言模型增强的基于时间注意力的推荐算法

为了解决LLM增强的推荐模型难以充分挖掘

ID所代表的实际语义和传统的时间序列模型无法应对用户兴趣变化较快场景的问题,本文提出了一个时间增强大语言模型的推荐算法。该算法通过特征提取模块生成用户的历史交互的协同特征以及具体交互时间的时间特征,进而提取其深层特征。此外,采用多样化的提示组合,从LLM中生成多个推荐结果,通过融合大语言模型与传统推荐模型的输出,可以获得更准确的推荐结果。

1.1 推荐模型的协同特征提取

特征提取模块首先对用户和商品的ID进行独热编码,同时使用时间上下文编码^[11]对用户商品的交互时间进行编码,在嵌入层将稀疏的编码映射为稠密的潜在向量。利用特征提取层构建了多层感知器(MLP),在嵌入向量中学习协同特征。

对于每个用户 $u \in U$ 、项目 $i \in I$,在嵌入层通过独热编码与嵌入矩阵相乘,生成用户嵌入向量 p_u 和项目嵌入向量 q_i 。

为了提取用户与项目交互的协同信息,将用户项目的嵌入向量连接后输入到MLP网络 $f(\cdot)$ 中,输出得到包含用户项目协同信息的向量 $\hat{e}_{u,i}$ 。采用双层MLP网络,并根据经验实现了塔式结构,将每一层的维度减半,如图1所示。

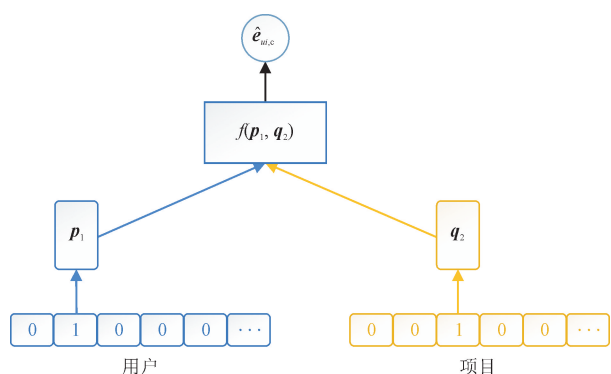


图1 协同特征提取

Fig. 1 Extraction of collaborative feature

1.2 推荐模型的时间特征提取

仅提取协同信息无法灵活地响应用户兴趣的动态变化,即处于不同时期的交互项目理应对用户产生不同的影响,提出了一个时间特征提取网络,用以

提取不同历史互动项目对当前项目的不同影响。

由于用户的偏好具有周期性,在推荐系统中,这意味着用户的行为和选择会随着时间的推移而重复出现,识别这种模式可以提高推荐的准确性和相关性,使推荐符合用户兴趣的周期性。受文献[12]中位置编码算法的启发,设计了时间嵌入方法,即

$$\tau(k) = \begin{cases} \sin\left(\frac{t_1}{10\,000}\right)^{2k/d}, & k \text{ 为偶数} \\ \cos\left(\frac{t_1}{10\,000}\right)^{2(k-1)/d}, & k \text{ 为奇数} \end{cases} \quad (1)$$

式中: $\tau(k)$ 为时间嵌入向量分量; k 为特定的维度; d 为整个嵌入的维度; t_1 为时间戳。式(1)能够自然地引导模型学习基于相对位置的注意力机制,模型在后续的时空注意力计算过程中,可以更高效地捕捉时间相关的信息,从而提升对时间动态的感知能力。

模型不同部分能够在最终预测中发挥不同程度的作用,是注意力机制的核心思想^[13]。在时间注意力的场景中,依据交互发生时间,为每个历史交互项目分配个性化权重,从而影响最终的预测结果。对于目标用户 u 和目标项目 i ,用户的历史交互项目 j 产生的时间信息权重 a_{uj} 为

$$a_{uj} = \frac{\exp(f(\tau_i, \tau_j))}{\sum_{j \in R_u \setminus i} \exp(f(\tau_i, \tau_j))} \quad (2)$$

时间特征提取如图 2 所示,将目标项目和历史交互项目的时间嵌入向量连接到一起输入到 MLP 网络中,输出得到包含用户项目交互时间信息的向量

$$\hat{e}_{ui,t} = \frac{1}{|R_u|} \sum_{j \in R_u \setminus i} f(\tau_i, a_{uj} \tau_j) \quad (3)$$

式中: R_u 为用户 u 的交互项目集合; $f(\cdot)$ 表示 MLP 网络。

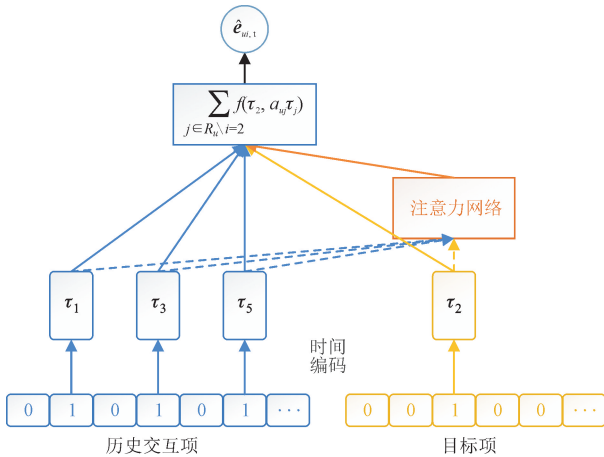


图 2 时间特征提取

Fig. 2 Extraction of time feature

1.3 信息融合

融合向量 $\hat{e}_{ui}^c, \hat{e}_{ui}^t$ 得到传统模型结合协同信息和历史交互时间信息的预测向量,得到传统推荐模型的预测向量,如图 3 所示。将预测向量输入到一个全连接层,以获得传统推荐模型的预测评分 $\hat{r}_{ui,R}$

$$\hat{e}_{ui,R} = \hat{e}_{ui}^c \parallel \hat{e}_{ui}^t \quad (4)$$

$$\hat{r}_{ui,R} = r_{\max} \sigma(W_1 \hat{e}_{ui,R}) \quad (5)$$

式中: $\hat{e}_{ui,R}$ 为预测向量; $\parallel$ 表示连接操作; W_1 为将向量转换为全连接层向量; $r_{\max}$ 为最大评分值。对于激活函数 σ , 选择 ReLU 函数,其能对稀疏的激活进行编码,非常适合稀疏的数据,并使模型不被过分放大^[14]。

传统推荐算法的损失函数是将均方根误差 (RMSE) 和平均绝对误差 (MAE) 结合,作为模型训练的损失函数,即

$$L_R = \sum_{u,i} [(\hat{r}_{ui} - r_{ui})^2 + |\hat{r}_{ui} - r_{ui}|] \quad (6)$$

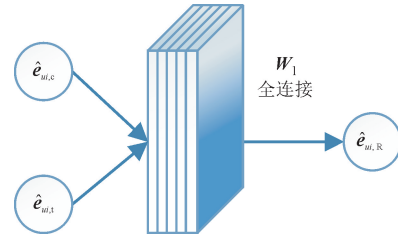


图 3 信息融合

Fig. 3 Fusion of information

1.4 大语言模型的推荐

传统推荐模型一般是将特征转换成相应的向量,忽视了与特征相关的语义信息,并且很难有效纳入辅助信息,这限制了传统模型的性能。根据具体评级预测任务的特点,针对具体评级预测任务,设计了一系列提示词作为大语言模型 LLM 的输入,以生成推荐结果。提示模板中,指令部分要求模型基于用户评分历史,综合考虑用户偏好、过往评分、电影特征及上映时间、导演、编剧、主演等情境因素,预测用户对某部电影的评分,并给出详细推理说明;信息部分提供了用户评分历史、用户特征(性别、职业、邮编、年龄等)以及电影详情(名称、上映年份、类型、用户最近评分时间等);输出格式示例明确了预测结果应包含电影标题、预测评分及评分理由;最终答案为 LLM 的输出内容 $\hat{r}_{ui,L}$ 。受文献[15]的启发,采用多样化的提示词来提高推荐模型性能。利用电影详情部分引出不同的推理路径,从而可以获得更可靠、更全面的结论。

1.5 对比学习

由于简单的使用大语言模型的输出 $\hat{r}_{u,L}$ 无法对模型参数进行训练调整,将大语言模型的预测评分编码为增强向量与传统推荐向量进行结合,从而对整个模型的可学习参数进行调整。

通过一个全连接层对大语言模型的预测评分进行编码,即

$$\hat{e}_{u,L} = W_2 \hat{r}_{u,L} \quad (7)$$

式中: W_2 为将大语言模型预测评分进行编码的可学习参数矩阵。传统推荐模型通常依赖用户-物品交互数据,存在冷启动、难以理解文本语义、泛化能力弱、难以融合多模态信息等问题,且推荐结果缺乏可解释性。而大语言模型凭借强大的自然语言理解能力,可以基于用户描述生成个性化推荐,无需大量历史数据,同时能够融合文本、图像等多模态信息,支持多轮交互优化,并提供清晰的推荐理由,从而有效弥补传统模型的不足。结合文献[16]中对比学习的思想,将大语言模型向量作为增强向量与传统推

荐向量进行对比,对比损失 L_C 可以看作是包含传统模型信息与大语言模型信息嵌入的组合,即

$$L_C = \sum_{u,i} -\lg \frac{\exp(\hat{e}_{u,R} \hat{e}_{u,L} / \tau)}{\sum_{(u,i)} \exp(\hat{e}_{u,R} \hat{e}_{u,L} / \tau)} \quad (8)$$

式中: τ 为温度超参数。

1.6 损失函数

由于推荐系统的主要目标是模拟用户与项目之间的交互,因此将所提对比学习损失视为补充,并利用多任务学习策略来联合训练传统的评分损失和对比损失,即

$$L = L_R + \lambda_1 L_C + \lambda_2 \theta \quad (9)$$

式中: λ_1 、 λ_2 分别为控制目标权重和正则化项的超参数; θ 为模型参数集。

这种损失函数将基于 LLM 的推荐模型和传统推荐模型的结果结合起来,能根据用户的历史交互行为和大语言模型根据多样化的提示词获得推荐评分,调整基于 LLM 的模型和传统推荐模型的贡献,整体框架如图 4 所示。

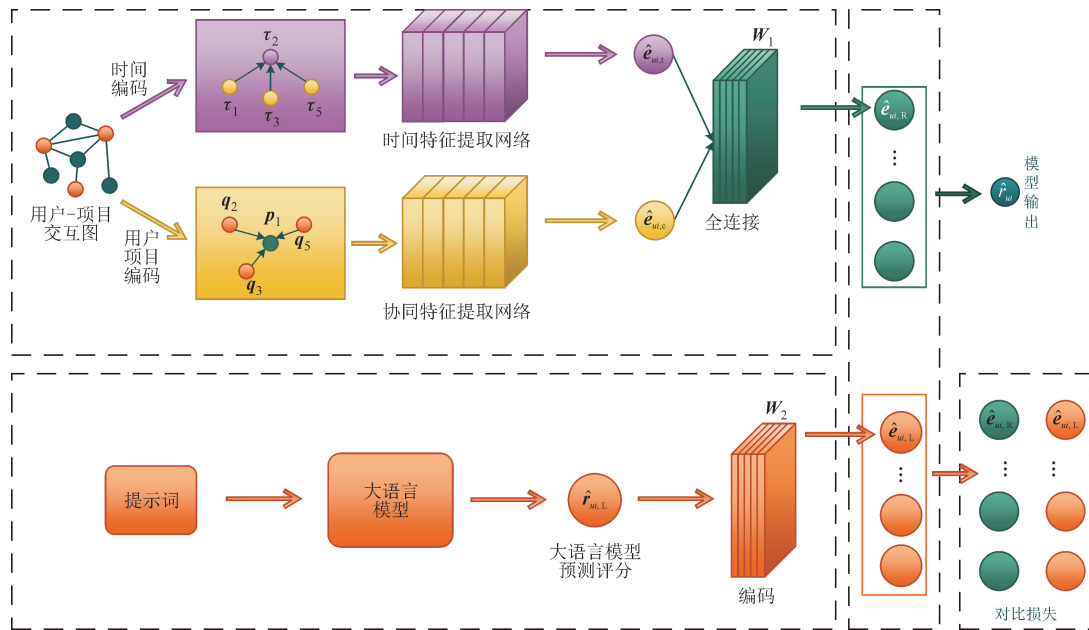


图 4 LLATR 整体框架

Fig. 4 Overall framework of LLATR

2 实验结果与分析

2.1 实验设置

本文算法的实验操作系统为 Ubuntu20.04.5, 显卡为英伟达 RTX3090。以联合损失作为损失函数,并使用随机梯度下降法^[17]作为优化算法。为了评估所提 LLATR 的有效性,在 MovieLens-100 K

(ML)^[18]和 Kaggle-Movie(KM)^[19]数据集上进行实验,这 2 个数据集广泛用于评估电影分级预测中的推荐算法。

在评分预测类推荐任务中,模型的预测结果通常以连续数值形式给出,故需使用回归类指标来衡量预测值与真实评分之间的偏差。采用均方根误差 (RMSE)、平均绝对误差 (MAE) 两项常用的误差指

标^[20],均方根误差、平均绝对误差计算式为

$$E_1 = \sqrt{\frac{1}{n} \sum_{u,i} (\hat{r}_{ui} - r_{ui})^2} \quad (10)$$

$$E_2 = \frac{1}{n} \sum_{u,i} |\hat{r}_{ui} - r_{ui}| \quad (11)$$

式中: n 为评分总数。

RMSE 对大误差更敏感,可用于惩罚极端偏差。相较于 RMSE,MAE 对异常值的敏感度较低,更能反映整体的平均误差水平。两者同时使用,可以兼顾整体预测精度与异常评分控制能力,从而更全面评估推荐算法的性能。

将 LLATR 与 CCPM^[21]、DeepFM^[22]、NCF^[23]、DCN^[24]、AFM^[25]、xDeepFM^[26]、AutoInt^[27]、LAMAR-DeepFM^[28] 等用于预测推荐系统评分的推荐模型相结合,为了使研究结果具有充分的再现性,使用开源的 LLaMA^[29] 模型作为 LLM 主要模型。

为确保公平性,采用基线原始论文中报告的最佳超参数设置,并使用网格搜索,对所有基线超参数进行微调。采用留一法策略^[23],对每个用户使用最后 1 个交互项进行测试,用倒数第 2 个交互项进行验证,其他交互项进行训练。在默认设置中,超参数 λ_1 为 0.1, λ_2 为 10^{-4} ,输入向量维度 d 为 64。

2.2 对比实验

在两个数据集上对基线方法和所提模型进行实验评估,以全面分析其在各种场景下的表现。各模型在这些数据集上的实验结果如表 1 所示,可知 LLATR 能够显著提升推荐效果。传统模型擅长挖掘协作信息,而 LLM 提供丰富的语义理解,两者的结合使得 LLATR 评分预测更加精准。

表 1 对比实验结果

Table 1 Result of comparative experiments

模型	E_1		E_2	
	KM	ML	KM	ML
DeepFM	0.987 3	1.120 4	0.776 5	0.872 9
NCF	1.037 9	1.058 1	0.780 6	0.832 3
DCN	1.029 9	1.112 1	0.779 4	0.859 1
AFM	1.037 8	1.093 8	0.801 6	0.838 5
xDeepFM	1.062 5	1.252 8	0.813 4	0.976 6
AutoInt	0.988 1	1.097 0	0.763 6	0.858 1
LAMAR-DeepFM	0.980 1	1.077 8	0.754 6	0.839 9
LLATR	0.960 2	1.023 6	0.732 9	0.822 7

DeepFM、NCF、DCN、AFM 和 xDeepFM 等传统深度推荐模型主要依赖于用户-物品交互数据来

学习协同信息。其中,DeepFM 结合了因子分解机和深度神经网络,在多个数据集上取得了较为稳定的表现,但 RMSE、MAE 仍然高于 LLATR 的。NCF、DCN 由于使用非线性变换来建模交互特征,在部分数据上有所提升,但整体误差依然较大。此外,AFM 通过注意力机制增强了特征交互建模能力,但其 RMSE、MAE 在两个数据集上均未能达到最优。xDeepFM 在 ML 上的 RMSE 最高,说明其在该数据集上的泛化能力较弱。

AutoInt 模型通过引入自注意力机制增强了特征建模能力,相较于传统模型,在 RMSE、MAE 方面均有一定的提升,表明其在某些数据集上能够更好地捕捉用户偏好。然而,尽管这些模型对推荐性能有所优化,但仍然未能在所有情况下达到最佳性能。

相比于 LAMAR 模型,LLATR 模型进一步利用时间信息,使得推荐更加符合用户的时序行为规律,提高了评分预测的准确性。这进一步验证了时间信息在 LLM 增强推荐系统中的重要性,为未来的研究和实践提供了重要参考。

LLATR 模型结合 LLM 提供的语义信息和传统推荐模型的协同过滤能力,实现了更精准的评分预测。从表 1 可以看出,LLATR 模型在 KM、ML 数据集上均取得了最低的 RMSE、MAE,表明其在预测用户评分方面更为准确。这主要归功于如下方面:①语义信息的增强作用,LLM 可以提供用户和物品的深层语义信息,使模型能够理解评分背后的潜在模式,而不仅仅依赖于协同信号;②协同与语义的互补性,传统模型依赖用户历史行为来建模偏好,而 LLM 能够补充文本、内容等外部信息,使推荐结果更具解释性和泛化能力;③时间信息的建模能力,相比于 LAMAR 模型,LLATR 模型进一步引入了时间信息的建模,LAMAR 作为一种利用 LLM 增强传统推荐模型的方法,虽然能提升推荐效果,但未能充分考虑时间因素,无法捕捉用户兴趣随时间变化的动态特性,而 LLATR 模型通过结合时间上、下文信息,使得模型在不同时间段下的推荐更加精准和个性化,提高了整体预测的可靠性;④模型融合策略的优化,LLATR 模型在融合协同和语义信息方面采用了有效的策略,使其能够充分发挥两种方法的优势,从而在不同数据集上都能取得稳定的提升。

为进一步验证所提方法在实际推荐中的有效性,以数据集中一个用户样本为例,对比分析传统模型与 LLATR 在评分预测过程中的差异,该用户近期多次观看了剧情类与传记类电影,并给予高评分。

针对同类型但风格更冷门的电影,传统模型由于缺乏上、下文理解能力,仅依赖标签与评分矩阵,未能准确捕捉到与用户偏好的深层语义关联,预测评分偏低,而 LLATR 模型借助大语言模型对该电影的导演、主角及历史背景的自动提取与语义建模,识别出与用户已评分影片的风格趋同性和人物类型偏好,预测评分达到 4.3,明显更贴近用户实际喜好。此外,从时间动态特性来看,该用户在过去 3 个月集中评分了多部基于真实人物的电影,显示出短期兴趣偏移趋势。LLATR 在模型中引入交互时间的时间注意力机制,使得近期行为在推荐权重中占据主导,从而对上述偏好变化做出及时响应。而对比模型未显式建模时间因素,导致推荐结果更多受历史平均偏好主导,无法及时跟进用户兴趣的变化,体现了 LLATR 在语义增强与时间敏感建模方面的显著优势,进一步验证了在应对个性化动态推荐任务中的有效性、独特性。

2.3 消融实验

为深入分析 LLATR 各个组件的贡献,进行了消融实验,构建了 3 个变体模型并进行实验,实验结果如表 2 所示。

表 2 消融实验结果

Table 2 Result of ablation experiments

模型	E_1		E_2	
	KM	ML	KM	ML
LLATR	0.970 3	1.045 7	0.738 6	0.824 7
w/o LLM	0.983 4	1.080 3	0.759 5	0.843 4
w/o Time	0.978 9	1.075 4	0.751 6	0.834 5
w/o Traditional	1.030 3	1.187 4	0.744 9	0.900 7

去除 LLM 语义信息后,模型的 RMSE、MAE 均有所上升,说明 LLM 在提升推荐准确性方面发挥了重要作用。传统推荐模型主要依赖用户-物品交互数据,无法有效利用文本、内容等外部信息,而 LLM 提供的丰富语义信息能够弥补这一不足,使推荐更加精准。时间信息的去除导致推荐性能下降,表明时间因素在 LLATR 中起到了关键作用。用户的兴趣随时间变化,而不考虑时间信息的模型容易受历史行为的干扰,无法准确捕捉用户的动态偏好。引入时间信息可以帮助模型更好地理解用户在不同时间段的兴趣偏好,提高推荐的时效性和准确性。去除传统推荐模型后,LLATR 的 RMSE、MAE 显著上升,说明协同信息对于推荐效果至关重要。尽管 LLM 具有强大的语义理解能力,但仅依靠 LLM

进行推荐的效果不如结合传统推荐模型。传统推荐模型能够有效挖掘用户-物品交互模式,而 LLM 主要提供文本语义增强,二者的结合才能实现最优推荐性能。

消融实验结果进一步验证了 LLATR 各个组件的必要性。LLM 语义增强能够提升推荐的泛化能力,时间信息增强有助于建模用户兴趣的动态变化,而传统推荐模型仍然是核心的协同信息来源。只有同时结合这 3 方面信息,才能实现最佳的推荐效果。

2.4 稀疏性适应能力分析

推荐系统中普遍存在数据稀疏问题,即大多数用户只与极少数物品发生过交互,导致传统协同过滤类模型难以有效建立用户偏好。在稀疏场景下,模型不仅难以获得充分的训练信号,而且容易陷入过拟合。为了验证所提 LLATR 模型在稀疏数据环境下的鲁棒性与适应能力,设计了稀疏度控制实验。

在原始 KM 数据集基础上,分别按照保留 30%、50%、70% 的用户-物品交互比例构造子集,模拟不同稀疏程度的推荐场景。保持其他实验设置不变,分别在这 3 个子集上测试 LLATR 与其他代表性模型的性能变化,实验结果如表 3 所示。

表 3 稀疏实验结果

Table 3 Result of sparse experiments

稀疏比例/%	模型	E_1	E_2
30	DeepFM	1.124 8	0.874 1
	AutoInt	1.105 5	0.860 3
	LAMAR-DeepFM	1.091 2	0.842 5
	LLATR (本文)	1.062 1	0.821 2
50	DeepFM	1.047 2	0.821 8
	AutoInt	1.027 8	0.804 2
	LAMAR-DeepFM	1.013 6	0.783 7
	LLATR (本文)	0.998 4	0.764 3
70	DeepFM	1.018 5	0.791 3
	AutoInt	1.016 1	0.786 9
	LAMAR-DeepFM	0.998 7	0.767 3
	LLATR (本文)	0.990 3	0.751 8

由表 3 可知,随着数据变得稀疏,各模型均出现一定程度的性能下降,但 LLATR 的 RMSE、MAE 上升幅度最小,展现出良好的鲁棒性。在用户交互数据有限的情况下,LLM 提供了额外的文本语义信息,有效缓解了行为稀疏带来的特征学习瓶颈,对比学习增强了模型对有限样本的判别能力。实验结果

表明,LLATR 具备较强的抗稀疏性能力,适用于真实应用中常见的长尾分布与冷启动场景。

2.5 超参数实验

在所提多任务损失函数中,超参数 λ_1 、 λ_2 分别用于平衡评分预测误差项与对比学习项之间的权重关系,以及控制模型参数正则化项的强度。其中, λ_1 控制传统评分预测损失与大语言模型语义增强向量之间对比损失的相对权重。若该值过小,可能导致模型过度依赖传统评分误差,忽视语义增强信息;若过大,则会过度拟合对比关系,弱化评分预测的主任务。当 $\lambda_2 = 1 \times 10^{-4}$ 时^[22-23], $\lambda_1 \in \{0.05, 0.1, 0.2, 0.5\}$, 选取最优值,使其在数据集 ML 上均表现出较佳的平衡性与鲁棒性。

当 λ_1 过小时,模型更多关注传统评分预测误差,忽略对比学习损失项的优化。推荐系统基本退化为传统协同过滤模型,仅依赖用户-项目交互数据,对大语言模型生成的深层语义信息利用不足。因此,模型难以有效捕捉用户偏好中的语义关联,导致预测误差偏大。当 λ_1 过大时模型对对比学习损失项的关注度过高。对比学习模块的权重过重,模型逐渐偏离评分预测主任务,转而优化语义对比损失。此时,模型可能会出现语义过拟合现象,即对大语言模型生成的语义特征高度依赖,但对评分预测任务的直接优化不足,导致 RMSE、MAE 增加。

在 ML 选取最优值, $\lambda_2 \in \{1 \times 10^{-5}, 5 \times 10^{-5}, 1 \times 10^{-4}, 5 \times 10^{-4}\}$ 时进行实验。当 λ_2 过小时,正则化项的抑制效果不足,模型倾向于对训练数据进行过度拟合。在训练集上的 RMSE、MAE 可能较低,但在测试集上,由于模型捕捉到大量噪声特征,泛化能力下降,测试集上的 RMSE、MAE 增加。随着 λ_2 的增加,正则化项有效抑制了模型参数的无限增长,保证了模型在训练和测试集上的平衡表现。模型能够有效建模协同信息与语义信息,同时控制参数规模,避免参数冗余或噪声学习,从而降低 RMSE、MAE。但当 λ_2 过大时,正则化项抑制作用增强,模型参数更新幅度受限。由于参数更新受限,模型难以捕捉复杂的语义关联与协同特征,模型表现出欠拟合倾向,RMSE、MAE 在训练集和测试集上均有所上升。

2.6 大模型影响

为探究 LLATR 算法对不同大语言模型的敏感性,在相同的推荐框架结构下,替换语义评分生成所使用的 LLM,对推荐性能进行横向对比分析。所选模型涵盖开源与闭源、英文与中文、多语言模型,具

体包括 GPT-3.5-Turbo(闭源,英文主导,API 调用)、LLaMA-2 7B(开源英文模型,结构紧凑)、ChatGLM-6B(开源中文优化模型)、BLOOMZ-7B(开源多语言模型)。

所有模型在提示词模板和输入数据保持一致条件下进行推理,输出的评分经统一编码方式参与对比学习与融合建模。实验在数据集 KM 上进行,结果如表 4 所示。

表 4 大模型影响

Table 4 Influence of LLM

模型	E_1	E_2
GPT-3.5-Turbo	0.964 5	0.729 4
LLaMA-2 7B	0.970 3	0.738 6
ChatGLM-6B	0.975 6	0.747 4
BLOOMZ-7B	0.982 7	0.753 9

从表 4 可以看出,GPT-3.5-Turbo 模型在 RMSE、MAE 上表现略优,得益于强大的上、下文理解与评分预测能力,尤其在理解提示信息和生成连贯合理评分解释方面表现突出。LLaMA-2 7B 模型的表现次之,虽然整体性能略逊,但作为开源模型,在可部署性、效率和稳定性方面具有较强优势,是综合权衡下的实际首选。ChatGLM-6B 模型对中文语义建模能力较强,但在评分任务中对复杂上下文处理略显不足。BLOOMZ-7B 作为多语言模型,泛化能力广,但生成内容缺乏一致性,导致在评分任务中的误差稍大。因此,LLATR 对 LLM 存在一定性能敏感性,不同模型输出的语义评分在质量和可训练性上存在差异,但所有模型均能提升推荐性能,验证了 LLATR 框架具备良好的跨模型兼容性与适配弹性。

2.7 推荐排序

推荐排序目的是向用户展示已排序的项目列表,为扩展本研究至排序任务,将原始评分数据集转换为隐式反馈数据:当用户对某个物品的评分大于或等于设定阈值 F 时,视为正反馈样本(1),否则视为负反馈样本(0)。由于原始评分为 1~5 分,设定 $F=3$ 来表示用户的偏好界限。

模型预测结果经 sigmoid 函数归一化为 $[0, 1]$, 排序概率计算式为

$$\hat{y}_{ui} = \text{sigmoid}(W_i \hat{e}_{ui}^R) \quad (12)$$

采用二元交叉熵损失函数,即

$$L_B = - \sum_{(u,i) \in R^+} \lg(\hat{y}_{ui}) - \sum_{(u,i) \in R^-} \lg(1 - \hat{y}_{ui}) \quad (13)$$

式中: R^+ 为正样本集合; R^- 为负样本集合。

为衡量模型在排序任务中的表现,引入如下评估指标: Precision@10(E_3), 在前 10 个推荐物品中, 正样本物品所占比例; NDCG@10(E_4), 归一化折损累计增益, 考虑推荐结果的相关性及其排序位置。在数据集 ML 上, 对 LLATR 与 DeepFM、AutoInt、LAMAR-DeepFM 等代表性模型进行了排序实验, 如表 5 所示。

表 5 排序性能

Table 5 Performance of ranking

模型	E_3	E_4
DeepFM	0.904	0.899
AutoInt	0.909	0.903
LAMAR-DeepFM	0.917	0.911
LLATR	0.932	0.925

由表 5 可知, 各模型在排序任务中整体表现良好, 其中, LLATR 在两个数据集上的两项指标均优于对比模型, 显示出其不仅具备良好的评分预测能力, 也能有效提升相关物品的排序质量。结果表明, 将大语言模型与时间特征、协同特征融合的方式, 在综合提升推荐效果方面具有显著优势。

3 结 论

本文提出了一种大语言模型增强的基于时间注意力的推荐框架(LLATR), 旨在优化推荐系统对用户兴趣变化的感知能力, 将时间信息与 LLM 结合, 以提高推荐系统的性能。LLATR 框架通过时间信息增强机制, 使 LLM 能够更好地理解用户兴趣的时间动态变化, 进而提升推荐的准确性和个性化水平, 得到如下主要结论。

(1) LLATR 框架融合了 LLM 强大的语义理解能力与用户-商品交互的时间信息。通过将时间维度纳入推荐系统, 能够更全面、精准地捕捉用户兴趣的动态变化, 从而显著提升了推荐系统的个性化和时效性, 为用户提供更符合当前需求和偏好的推荐内容。

(2) 设计的时间交互特征处理方法不仅考虑了用户与商品交互的时间先后顺序, 还深入挖掘了不同时间间隔内用户交互行为的潜在关联。通过这种方式, 增强了模型对用户兴趣变化的响应能力, 使其在时间敏感的推荐任务中, 能够迅速捕捉用户兴趣的转变, 提供更具针对性的推荐, 具有显著优势。

(3) 在两个类型、不同规模数据集上进行实验, 结果表明, LLATR 在均方根误差、平均绝对误差、准确率等多个关键指标上均超越了传统推荐方法和现有的 LLM 增强推荐模型。这充分证明了 LLATR 框架具有广泛的适用性, 可为不同领域的推荐系统提供有效的解决方案。

参考文献:

- [1] 刘华玲, 马俊, 张国祥. 基于深度学习的内容推荐算法研究综述 [J]. 计算机工程, 2021, 47(7): 1-12.
LIU Hualing, MA Jun, ZHANG Guoxiang. Review of studies on deep learning-based content recommendation algorithms [J]. Computer Engineering, 2021, 47(7): 1-12.
- [2] 王立才, 孟祥武, 张玉洁. 上下文感知推荐系统 [J]. 软件学报, 2012, 23(1): 1-20.
WANG Licai, MENG Xiangwu, ZHANG Yujie. Context-aware recommender systems [J]. Journal of Software, 2012, 23(1): 1-20.
- [3] 陈海涵, 吴国栋, 李景霞, 等. 基于注意力机制的深度学习推荐研究进展 [J]. 计算机工程与科学, 2021, 43(2): 370-380.
CHEN Haihan, WU Guodong, LI Jingxia, et al. Research advances on deep learning recommendation based on attention mechanism [J]. Computer Engineering & Science, 2021, 43(2): 370-380.
- [4] 秦继伟, 郑庆华, 郑德立, 等. 结合评分和信任的协同推荐算法 [J]. 西安交通大学学报, 2013, 47(4): 100-104.
QIN Jiwei, ZHENG Qinghua, ZHENG Deli, et al. A collaborative recommendation algorithm based on ratings and trust [J]. Journal of Xi'an Jiaotong University, 2013, 47(4): 100-104.
- [5] 彭梓航, 张全贵, 金海波, 等. 基于时间块动态图神经网络的序列推荐方法 [J/OL]. 计算机应用研究. [2025-01-27]. <https://doi.org/10.19734/j.issn.1001-3695.2024.12.0518>.
PENG Zihang, ZHANG Quanguai, JIN Haibo, et al. Time-block-based dynamic graph neural network for sequential recommendation [J/OL]. Application Research of Computers. [2025-01-27]. <https://doi.org/10.19734/j.issn.1001-3695.2024.12.0518>.
- [6] GRAVES A. Supervised sequence labelling with recurrent neural networks [M]. Berlin, Germany: Springer, 2012.
- [7] 吴文隆, 尹海莲, 王宁, 等. 大语言模型和知识图谱协同的跨域异质数据查询框架 [J]. 计算机研究与发

- 展, 2025, 62(3): 605-619.
- WU Wenlong, YIN Hailian, WANG Ning, et al. A synergetic LLM-KG framework for cross-domain heterogeneous data query [J]. *Journal of Computer Research and Development*, 2025, 62(3): 605-619.
- [8] 吕学强, 万甜, 马登豪, 等. LLM-PKE: 一种集成大模型与多特征网络的专利关键词提取方法研究 [J/OL]. *数据分析与知识发现*. [2025-02-19]. <http://kns.cnki.net/kcms/detail/10.1478.G2.20250219.1348.006.html>.
- LV Xueqiang, WAN Tian, MA Denghao, et al. LLM-PKE: research on a patent keywords extraction method integrating large language models and multi-feature networks [J/OL]. *Data Analysis and Knowledge Discovery*. [2025-02-19]. <http://kns.cnki.net/kcms/detail/10.1478.G2.20250219.1348.006.html>.
- [9] WANG Changdong, XI Wudong, HUANG Ling, et al. A BP neural network based recommender framework with attention mechanism [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2022, 34(7): 3029-3043.
- [10] 黄文, 李振江. 传统模式与大语言模型下推荐系统的比较研究 [J]. *软件导刊*, 2025, 24(2): 204-210.
- HUANG Wen, LI Zhenjiang. Recommendation systems: a comparative study of traditional models and pre-trained model approaches [J]. *Software Guide*, 2025, 24(2): 204-210.
- [11] HUANG Chao, XU Huance, XU Yong, et al. Knowledge-aware coupled graph neural network for social recommendation [C]//*Proceedings of the AAAI Conference on Artificial Intelligence*. Palo Alto, CA, USA: AAAI Press, 2021: 4115-4122.
- [12] KANG Wangcheng, MCAULEY J. Self-attentive sequential recommendation [C]//*2018 IEEE International Conference on Data Mining*. Piscataway, NJ, USA: IEEE, 2018: 197-206.
- [13] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate [EB/OL]. [2025-02-16]. <https://arxiv.org/abs/1409.0473>.
- [14] GLOROT X, BORDES A, BENGIO Y. Deep sparse rectifier neural networks [C]//*Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*. Chia Laguna Resort, Sardinia, Italy: PMLR, 2011: 315-323.
- [15] DENG Sinuo, WU Lifang, SHI Ge, et al. Learning to compose diversified prompts for image emotion classification [J]. *Computational Visual Media*, 2024, 10(6): 1169-1183.
- [16] LIN Zihan, TIAN Changxin, HOU Yupeng, et al. Improving graph collaborative filtering with neighborhood-enriched contrastive learning [C]//*Proceedings of the ACM Web Conference 2022*. New York, USA: ACM, 2022: 2320-2329.
- [17] 史加荣, 王丹, 尚凡华, 等. 随机梯度下降算法研究进展 [J]. *自动化学报*, 2021, 47(9): 2103-2119.
- SHI Jiarong, WANG Dan, SHANG Fanhua, et al. Research advances on stochastic gradient descent algorithms [J]. *Acta Automatica Sinica*, 2021, 47(9): 2103-2119.
- [18] HARPER F M, KONSTAN J A. The MovieLens datasets: history and context [J]. *ACM Transactions on Interactive Intelligent Systems*, 2015, 5(4): 19.
- [19] LUO Sichun, ZHANG Xinyi, XIAO Yuanzhang, et al. HySAGE: a hybrid static and adaptive graph embedding network for context-drifting recommendations [C]//*Proceedings of the 31st ACM International Conference on Information & Knowledge Management*. New York, USA: ACM, 2022: 1389-1398.
- [20] XIONG Fei, SUN Haoran, LUO Guixun, et al. Graph attention network with high-order neighbor information propagation for social recommendation [C]//*Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence (IJCAI-24)*. California, USA: IJCAI, 2024: 2478-2486.
- [21] LIU Qiang, YU Feng, WU Shu, et al. A convolutional click prediction model [C]//*Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. New York, USA: ACM, 2015: 1743-1746.
- [22] GUO Huifeng, TANG Ruiming, YE Yunming, et al. DeepFM: a factorization-machine based neural network for CTR prediction [C]//*Proceedings of the 26th International Joint Conference on Artificial Intelligence*. Palo Alto, CA, USA: AAAI Press, 2017: 1725-1731.
- [23] HE Xiangnan, LIAO Lizi, ZHANG Hanwang, et al. Neural collaborative filtering [C]//*Proceedings of the 26th International Conference on World Wide Web*. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee, 2017: 173-182.
- [24] WANG Ruoxi, FU Bin, FU Gang, et al. Deep & cross network for ad click predictions [C]//*Proceedings of the ADKDD'17*. New York, USA: ACM, 2017: 12.
- [25] XIAO Jun, YE Hao, HE Xiangnan, et al. Attentional

- factorization machines; learning the weight of feature interactions via attention networks [C]//Proceedings of the 26th International Joint Conference on Artificial Intelligence. Palo Alto, CA, USA; AAAI Press, 2017: 3119-3125.
- [26] LIAN Jianxun, ZHOU Xiaohuan, ZHANG Fuzheng, et al. xDeepFM: combining explicit and implicit feature interactions for recommender systems [C]//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. New York, USA: ACM, 2018: 1754-1763.
- [27] SONG Weiping, SHI Chence, XIAO Zhiping, et al. AutoInt: automatic feature interaction learning via self-attentive neural networks [C]//Proceedings of the 28th ACM International Conference on Information and Knowledge Management. New York, USA: ACM, 2019: 1161-1170.
- [28] LUO Sichun, WANG Jiansheng, ZHOU Aojun, et al. Large language models augmented rating prediction in recommender system [C]//ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing. Piscataway, NJ, USA: IEEE, 2024: 7960-7964.
- [29] TOUVRON H, LAVRIL T, IZACARD G, et al. Llama: open and efficient foundation language models [EB/OL]. [2025-02-10]. <https://arxiv.org/abs/2302.13971>.
- (编辑 赵炜)