Advances in Applied Mathematics 应用数学进展, 2026, 15(6), 263-273
Published Online June 2026 in Hans. https://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2026.156285

Hans汉斯

基于大语言模型强化的个性化推荐算法研究

贾佳奇，周文学*

兰州交通大学数理学院，甘肃 兰州

收稿日期：2026年5月18日；录用日期：2026年6月11日；发布日期：2026年6月18日

摘 要

随着互联网平台内容规模的持续扩张，用户在海量候选项中快速找到符合自身兴趣的信息变得越来越困难，个性化推荐成为提升用户体验的重要手段。现有推荐方法大多以用户－项目交互数据为核心进行偏好建模，但在用户兴趣动态变化、交互数据稀疏以及多源异构信息利用不足等问题上仍存在明显局限。针对上述问题，本文提出了一种结合大语言模型语义增强、知识图谱结构建模与深度强化学习决策的个性化推荐方法。该方法将推荐过程建模为序列决策问题，在用户侧结合行为序列表示与语义画像构建状态表示，在项目侧联合学习文本语义表示与知识图谱结构表示，并通过深度Q网络完成候选项目的价值评估与推荐决策。在Yelp和Amazon-Book两个公开数据集上进行了实验，并与多种代表性推荐方法进行了对比。结果表明，所提出的方法在HR@K以及NDCG@K等指标上均取得了较优表现，验证了所提框架的有效性。

关键词

个性化推荐，深度强化学习，大语言模型，跨视图表示学习，知识图谱

Research on Personalized Recommendation Algorithm Enhanced by Large Language Models

Jiaqi Jia, Wenxue Zhou*

School of Mathematics and Physics, Lanzhou Jiaotong University, Lanzhou Gansu

Received: May 18, 2026; accepted: June 11, 2026; published: June 18, 2026

Abstract

With the continuous expansion of the content scale on internet platforms, it has become increas-

*通讯作者。

文章引用: 贾佳奇, 周文学. 基于大语言模型强化的个性化推荐算法研究[J]. 应用数学进展, 2026, 15(6): 263-273.
DOI: 10.12677/aam.2026.156285

ingly difficult for users to quickly find information that matches their interests from the vast number of options. Personalized recommendations have become an important means to enhance user experience. Most existing recommendation methods mainly model preferences based on user-item interaction data. However, there are still significant limitations in addressing issues such as dynamic changes in user interests, sparse interaction data, and insufficient utilization of multi-source heterogeneous information. To address these problems, this paper proposes a personalized recommendation method that combines semantic enhancement of large language models, knowledge graph structure modeling, and deep reinforcement learning decision-making. This method models the recommendation process as a sequence decision-making problem. On the user side, it combines behavioral sequence representation and semantic profiling to construct state representation. On the item side, it jointly learns text semantic representation and knowledge graph structure representation, and completes the value assessment and recommendation decision through a deep Q-network. Experiments were conducted on two public datasets, Yelp and Amazon-Book, and compared with several representative recommendation methods. The results show that the proposed method achieves superior performance in metrics such as HR@K and NDCG@K, verifying the effectiveness of the proposed framework.

Keywords

Personalized Recommendation, Deep Reinforcement Learning, Large Language Model, Cross-View Representation Learning, Knowledge Graph

1. 绪论

在信息过载日益加剧的互联网环境下，个性化推荐系统已成为提升用户体验与平台效率的关键技术。目前多数推荐方法依赖用户与项目的静态交互数据进行偏好建模，虽在特定场景下取得一定效果，但在用户兴趣动态变化、交互数据稀疏以及多源异构信息利用不足等方面仍存在明显局限。特别是评论文本、项目描述等非结构化语义信息，以及知识图谱所蕴含的实体关系与结构关联，尚未在统一决策框架中得到充分利用。针对上述问题，本文提出了一种融合大语言模型语义增强、知识图谱结构建模与深度强化学习决策的个性化推荐方法，将推荐过程建模为序列决策问题，通过整合用户行为序列、语义画像与项目结构特征，提升模型对用户当前偏好与候选项目特征的刻画能力，从而实现更精准、稳定的推荐。

2. 基于大语言模型与强化学习的个性化推荐算法模型

2.1. 模型框架

模型整体架构如图 1 所示，其运行流程如下：首先，利用大语言模型对用户评论与项目描述进行语义抽取，生成用户语义画像与项目语义画像。其次，在用户侧，采用门控循环单元与多头自注意力机制对历史交互序列进行编码，并与用户语义画像融合，形成当前状态表示。在项目侧，基于知识图谱嵌入与关系感知注意力传播学习项目的结构表示，并通过跨视图对比学习方法将项目语义表示与结构表示进行对齐与融合，获得最终的动作表示。最后，将状态与动作表示输入深度 Q 网络，通过 Dueling 和 Double Q 机制进行价值评估与推荐决策。

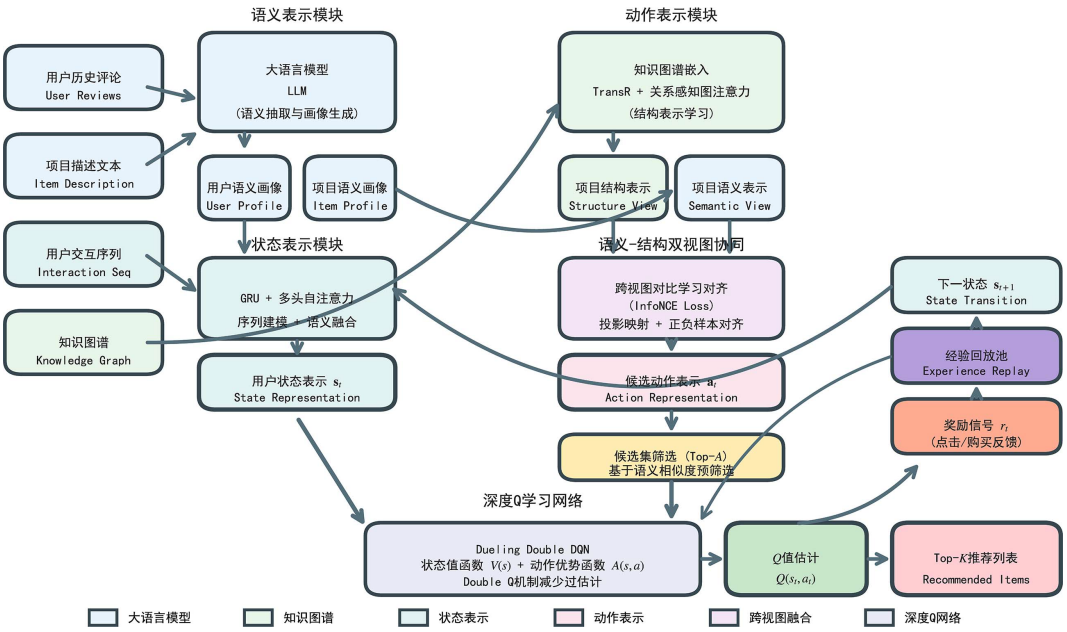

Figure 1. Model framework figure
图 1. 模型框架图

2.2. 语义表示模块：偏好与内容的深度理解

该模块旨在从非结构化文本中提取高层次的偏好与内容特征，以弥补交互数据的稀疏性。我们采用大语言模型通过提示词工程分别构建用户和项目的语义画像[1]。用户语义画像构建：对于每个用户 u，设其历史评论集合为：$C_u=\{c_1^u,c_2^u,\cdots,c_m^u\}$，输入至大语言模型。通过大语言模型提示词工程，引导模型总结出用户的偏好维度模板，即为用户语义画像：e_u^{sem}。项目语义画像构建：类似地，对于每个项目 i，我们会收集它的标题、简介、描述、类别文本等内容，组成项目文本信息集合：$D_i=\{Title_i,Desc_i,Meta_i\}$，参照用户画像的构建思路，这里同样借助大语言模型，通过提示词模板对项目文本进行总结，提取出项目的适配人群描述模板，然后生成项目语义画像 e_i^{sem}，如图 2 所示。

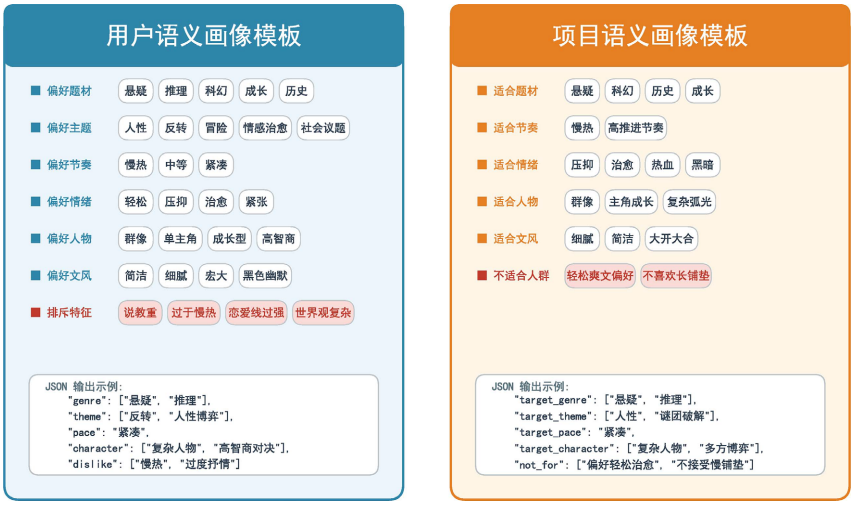

Figure 2. Example of generating semantic profiles for users and projects
图 2. 用户、项目语义画像生成示例

2.3. 状态表示模块：动态兴趣的捕捉与融合

状态表示模块，把 GRU 的时序建模能力[2]、多头自注意力的关键行为识别能力[3]，还有用户语义画像对长期偏好的刻画能力结合在一起。通过这种融合方式，模块能更全面地表达用户的当前状态。它一方面保留了用户长期稳定的偏好特征，另一方面也捕捉到了近期的兴趣变化。

序列行为编码：设用户在时刻 t 交互项目的嵌入表示为 i_t，该表示可由项目侧表示学习模块提供，用于刻画历史交互项目的内容与结构特征。使用门控循环单元 GRU 对序列进行建模：$h_t = GRU\left(i_t, h_{t-1}\right)$。$h_t$ 是 t 时刻的隐藏状态。最终隐藏状态 h_t 包含了截止当前时刻的序列信息。

关键行为提取：并非序列中所有行为对当前状态贡献相同。因此引入多头自注意力机制来区分不同历史行为的重要性。将隐藏状态序列 $H = \left[h_1 \cdots h_t\right]$ 作为输入，通过线性变换得到查询、键、值矩阵，计算缩放点积注意力，通过多个注意力头并行计算并拼接结果，得到加权后的序列表示 e_u^{seq}。

状态融合：将加权后的序列表示 e_u^{seq} 与用户语义画像 e_u^{sem} 拼接，并通过线性变换压缩维度，形成最终的强化学习状态表示 s_u。该状态表示同时编码了用户的短期行为模式与长期稳定偏好。

2.4. 动作表示模块：语义与结构的双视图协同建模

动作表示模块把知识图谱[4]的结构信息和项目的语义信息放在一起建模，让候选项目在进入 Q 网络之前，就有了更丰富、更有区分度的多视图表示。

2.4.1. 结构视图：基于知识图谱的关系感知传播

本工作基于图注意力网络架构[5]，通过递归传播嵌入来捕获高阶结构信息，引入注意力机制动态学习邻居节点的重要性。对于项目 i，通过关系感知注意力机制计算并更新邻域信息表示 e_{N_h}；通过求和后进行非线性变换聚合实体及其邻域表示，生成新的嵌入：$f = \text{LeakyReLU}\left(W\left(e_h + e_{N_h}\right)\right)$。经过 L 层传播后，项目 i 获得多阶表示 $\left\{e_i^{(1)}, \cdots, e_i^{(L)}\right\}$，为融合不同阶数的连通性信息，使用层聚合机制聚合各层输出：$e_i^{str} = e_i^{(0)} + \cdots + e_i^{(L)}$。其中 L 为超参数，e_i^{str} 表示项目结构向量的嵌入表示。

2.4.2. 双视图协同表示：跨视图对比对齐

为融合项目语义信息与结构信息，本文设计了语义－结构双视图协同表示机制[6]。首先，将项目语义向量 e_i^{sem} 和结构向量 e_i^{str} 通过可学习投影矩阵映射至同一空间。构造跨视图对比学习正负样本对[7]：同一项目的 $\left(e_i^{sem}, e_i^{str}\right)$ 互为正样本，同批次中不同项目的视图互为负样本。采用 InfoNCE 损失函数进行双向对比对齐，定义语义到结构方向的损失 $L_{s\to t}$ 与结构到语义方向的损失 $L_{t\to s}$，最终跨视图对齐损失为 $L_{CL} = \frac{1}{2}\left(L_{s\to t} + L_{t\to s}\right)$。该方法在保留各视图特有信息的同时增强跨视图语义一致性。最后，将对齐后的语义向量与结构向量拼接并经线性映射压缩，得到最终动作表示 $a_i = f\left(\left[\tilde{e}_i^{str} \,\|\, \tilde{e}_i^{sem}\right]\right)$ 该机制有效缓解了多源特征融合时的视图偏移问题，提升了候选项目表示的稳定性与判别性。

2.5. 深度 Q 学习网络

深度 Q 学习网络负责评估候选项目的推荐价值[8]。首先基于相似度筛选 $t+1$ 个项目构成候选动作空间 A_t。网络采用 Dueling 结构[9]，将 Q 值分解为 $Q(s,a) = V(s) + A(s,a)$，并引入 Double-Q 机制[10]，目标值计算为 $y_i = r_{t+1} + \gamma Q'\left(s_{t+1}, \text{argmax}_a Q\left(S_{t+1}, a; \theta\right); \theta'\right)$，以缓解过估计问题。奖励函数设为 $r_t = \begin{cases} 1, a_t = i_{t+1} \\ 0, a_t \neq i_{t+1} \end{cases}$。

训练采用经验回放与目标网络软更新。整体损失函数为 $L = L_Q + \lambda L_{CL} + \beta L_{reg}$，联合优化 Q 值预测、跨视图对齐与正则化项，实现表示学习与决策优化的协同训练。

3. 实验设计与结果分析

3.1. 实验设置

3.1.1. 数据集

本文选取 Yelp 和 Amazon-Book 两个公开数据集进行实验。Yelp 数据集来自本地生活服务平台，包含用户对商户的评分、评论及商户属性信息；Amazon-Book 数据集来自亚马逊图书领域，包含用户评分、评论及图书类别、描述等元数据。两个数据集分别代表服务类与内容类推荐场景，有助于验证方法的泛化能力。

数据预处理包括：1) 将评分 ≥ 4 的交互视为正反馈；2) 采用 10-core 过滤，保留交互次数 ≥ 10 的用户和项目；3) 知识图谱中保留出现 ≥50 次的关系类型；4) 按时间顺序以 6:2:2 比例划分训练集、验证集和测试集。预处理后数据统计如表 1 所示。

Table 1. Data dimension display table
表 1. 数据维度展示表

		Amazon-Book	Yelp
用户项目交互	用户数	70,679	45,919
	项目数	24,915	45,538
	交互数	847,733	1,185,056
知识图谱	实体数	88,572	90,961
	关系数	39	42
	元组数	2,557,746	1,853,704

3.1.2. 基线模型与评价指标

本文选取四类基线模型：传统推荐(BPR-MF、NCF)、序列推荐(GRU4Rec、Caser、SASRec、BERT4Rec)、知识图谱增强推荐(RippleNet、KGAT)以及强化学习推荐(MBCAL)。采用 HR@K 和 NDCG@K (K = 5, 10, 20)作为评价指标。主要超参数设置如表 2 所示。

Table 2. Hyperparameter setting table
表 2. 超参数设置表

超参数	超参数含义	值
d	嵌入维度	128
d_h	GRU 隐藏层维度	128
H	多头注意力头数	4
L	知识图谱传播层数	2
Top-A	候选项目数目	1000
γ	折扣因子	0.7
learning rate	学习率	0.0005

续表

batch size	批大小	256
replay size	经验回放池大小	50000
τ	目标网络软更新系数	0.005
dropout	丢弃率	0.2
λ_1	跨视图对齐损失权重	0.1
λ_2	L_2 正则化系数	1×10^{-5}
epoch	训练轮数	100
optimizer	优化器	Adam

3.1.3. 大语言模型配置

本文采用 GPT-3.5-turbo 与 LLaMA-2-7B 两种大语言模型进行对比实验。其中，GPT-3.5-turbo 通过 OpenAI API 调用，LLaMA-2-7B 采用本地化部署。两种大语言模型的性能对比表明，GPT-3.5-turbo 生成的语义画像在 NDCG@10 指标上平均优于 LLaMA-2-7B 约 3.5%，但其单用户画像生成时间增加约 40%。在计算开销方面，语义画像模块占总训练时间的 32%，该部分为离线生成，其中大语言模型调用占主导地位。最终选取 GPT-3.5-turbo 作为研究所用的大语言模型。

用户侧提示词模板设计如下：“根据以下用户评论，总结该用户的偏好维度，包括喜欢的类别、风格、价格敏感度等，输出 JSON 格式。评论内容：{comment_text}”。项目侧提示词模板设计如下：“根据以下项目描述，生成该项目的适配人群画像，包括适合的用户类型、使用场景等，输出 JSON 格式。项目信息：{item_text}”。为评估提示词鲁棒性，本文设计了三种变体模板，分别为详细版、简洁版与示例引导版，并在 Amazon-Book 数据集上进行对比。实验结果显示，HR@10 的最大波动幅度为 2.1%，表明模型对提示词格式具有一定鲁棒性。

3.2. 实验结果与分析

3.2.1. 对比实验

表 3 和表 4 展示了各模型在两个数据集上的实验结果。

Table 3. Experimental results of the Amazon-Book dataset
表 3. Amazon-Book 数据集实验结果

	Amazon-Book 数据集					
模型	**HR@5**	**HR@10**	**HR@20**	**NDCG@5**	**NDCG@10**	**NDCG@20**
BPR-MF	0.0964	0.1468	0.2147	0.0689	0.0851	0.1022
NCF	0.1035	0.1549	0.2238	0.0736	0.0902	0.1076
GRU4Rec	0.1188	0.1716	0.2427	0.0854	0.1024	0.1203
Caser	0.1139	0.1662	0.2361	0.0812	0.0979	0.1156
SASRec	0.1314	0.1847	0.2573	0.0946	0.1117	0.1301
BERT4Rec	0.1385	0.1928	0.2664	**0.1061**	0.1236	0.1424
RippleNet	0.1213	0.1749	0.2472	0.0872	0.1044	0.1228
KGAT	0.1322	0.1861	0.2615	0.0951	0.1123	0.1310
MBCAL	0.1358	0.1905	0.2691	0.0974	0.1149	0.1349
本文模型	**0.1446**	**0.1987**	**0.2718**	0.1054	**0.1243**	**0.1436**

Table 4. Experimental results of the Yelp dataset
表 4. Yelp 数据集实验结果

	Yelp 数据集					
模型	**HR@5**	**HR@10**	**HR@20**	**NDCG@5**	**NDCG@10**	**NDCG@20**
BPR-MF	0.0827	0.1283	0.1931	0.0581	0.0727	0.0891
NCF	0.0889	0.1356	0.2015	0.0627	0.0776	0.0943
GRU4Rec	0.1026	0.1528	0.2237	0.0739	0.0900	0.1079
Caser	0.0983	0.1486	0.2189	0.0702	0.0863	0.1038
SASRec	0.1175	0.1698	0.2414	0.0856	0.1023	0.1204
BERT4Rec	0.1246	0.1779	0.2501	**0.0970**	0.1141	0.1322
RippleNet	0.1067	0.1578	0.2298	0.0768	0.0931	0.1113
KGAT	0.1163	0.1684	0.2425	0.0845	0.1012	0.1200
MBCAL	0.1208	0.1736	0.2517	0.0881	0.1050	0.1248
本文模型	**0.1294**	**0.1826**	**0.2543**	0.0962	**0.1153**	**0.1336**

从表 3 和表 4 可见，本文模型在大部分指标上均取得最优结果，验证了融合 LLM 语义增强、知识图谱结构传播与强化学习排序框架的有效性。

1) 与传统推荐模型对比：静态模型 BPR-MF 和 NCF 表现落后。本文模型在 Amazon-Book 上 HR@10 达 0.1987，较 BPR-MF (0.1468)提升 35.35%，较 NCF (0.1549)提升 28.28%；在 Yelp 上分别提升 42.32% 和 34.66%，说明序列建模与强化学习对动态兴趣捕捉的必要性。

2) 与序列推荐模型对比：SASRec 和 BERT4Rec 优于 GRU4Rec 和 Caser，表明自注意力机制的优势。BERT4Rec 在 NDCG@5 上略优于本文模型，体现其精排能力；但本文模型在 HR@5/10/20 及 NDCG@10/20 上全面超越。Amazon-Book 上 HR@10 由 0.1928 提至 0.1987，NDCG@10 由 0.1236 提至 0.1243；Yelp 上 HR@10 由 0.1779 提至 0.1826，NDCG@20 由 0.1322 提至 0.1336。说明仅靠序列行为表达偏好不足，文本语义与结构信息的增强有效提升了排序质量。

3) 与知识图谱推荐模型对比：KGAT 在 Amazon-Book 上 HR@20 达 0.2615，体现结构化信息价值。本文模型优于 KGAT，Amazon-Book 上 HR@10 (0.1987 vs 0.1861)和 NDCG@10 (0.1243 vs 0.1123)均更高，Yelp 上优势一致。表明结构信息需结合序列建模与动态决策才能更好响应即时兴趣。

4) 与强化学习推荐模型对比：MBCAL 在 Amazon-Book 上 HR@20 达 0.2691，覆盖能力较强。本文模型融合语义画像与结构增强后全面优于 MBCAL，Amazon-Book 上 HR@10 由 0.1905 提至 0.1987，NDCG@10 由 0.1149 提至 0.1243；Yelp 上 HR@10 由 0.1736 提至 0.1826，NDCG@10 由 0.1050 提至 0.1153。说明 RL 性能高度依赖状态与动作表示质量，多源融合机制有效增强了 Q 网络的价值估计能力。

5) 跨数据集差异分析：各模型在 Amazon-Book 上表现普遍优于 Yelp，原因包括：1) 图书语义信息稳定、文本质量高，利于语义画像构建；2) 图书间类别、作者等关系规整，便于 KG 传播；3) 图书消费兴趣主题连续性强，适合序列建模与 RL 排序。Yelp 场景受时间、地点、社交等情境因素影响大，兴趣切换频繁、噪声更多。该差异说明本文模型在语义结构稳定、兴趣演化连续的场景中优势更突出，也提示需在复杂生活服务场景中进一步引入情境信息与多行为反馈。

3.2.2. 消融实验

为分析各模块贡献，设计四种变体：w/o LLM (移除语义增强)、w/o KG (移除知识图谱)、w/o CL (移

除跨视图对齐)、w/o RL (移除强化学习)。实验结果如表 5、表 6 所示。

Table 5. Ablation experiment results of the Amazon-Book dataset
表 5. Amazon-Book 数据集消融实验结果

Amazon-Book 数据集						
模型变体	**HR@5**	**HR@10**	**HR@20**	**NDCG@5**	**NDCG@10**	**NDCG@20**
w/o LLM	0.1329	0.1862	0.2571	0.0963	0.1132	0.1310
w/o KG	0.1341	0.1875	0.2590	0.0972	0.1141	0.1320
w/o CL	0.1339	0.1865	0.2580	0.0965	0.1138	0.1315
w/o RL	0.1318	0.1851	0.2554	0.0954	0.1120	0.1298
Full Model	**0.1446**	**0.1987**	**0.2718**	**0.1054**	**0.1243**	**0.1436**

Table 6. Ablation experiment results of the Yelp dataset
表 6. Yelp 数据集消融实验结果

Amazon-Book 数据集						
模型变体	**HR@5**	**HR@10**	**HR@20**	**NDCG@5**	**NDCG@10**	**NDCG@20**
w/o LLM	0.1186	0.1711	0.2415	0.0884	0.1052	0.1231
w/o KG	0.1198	0.1724	0.2433	0.0892	0.1061	0.1243
w/o CL	0.1190	0.1719	0.2422	0.0890	0.1061	0.1239
w/o RL	0.1179	0.1702	0.2398	0.0878	0.1044	0.1219
Full Model	**0.1294**	**0.1826**	**0.2543**	**0.0962**	**0.1153**	**0.1336**

从表 4 可见，完整模型在所有指标上均达最优，四个核心模块均有正向贡献，但贡献程度存在差异。

1) 强化学习模块贡献最显著：移除 RL 后性能下降最大。Amazon-Book 上 HR@10 从 0.1987 降至 0.1851，NDCG@10 从 0.1243 降至 0.1120；Yelp 上 HR@10 从 0.1826 降至 0.1702，NDCG@10 从 0.1153 降至 0.1044。说明序列决策建模与 Q 值估计对捕捉用户兴趣演化至关重要。

2) 大语言模型语义增强模块作用重要：移除 LLM 后降幅接近去除 RL。Amazon-Book 上 HR@10 从 0.1987 降至 0.1862，NDCG@10 从 0.1243 降至 0.1132。图书场景中主题、作者等语义信息难以仅通过项目 ID 表达，验证了语义画像构建的必要性。

3) 知识图谱结构模块稳定增强项目表示：去除 KG 后性能稳定下降，说明显式关系信息有助于挖掘高阶结构语义。其贡献略低于 LLM 和 RL，但与语义信息互补，共同构成优质项目表征。

4) 跨视图对齐方法增益稳定：去除后性能下降幅度接近去除 LLM，说明缺少该模块时语义表示难以有效发挥作用。完整模型相比 w/o CL，Amazon-Book 上 HR@10 从 0.1865 提至 0.1987，NDCG@10 从 0.1138 提至 0.1243；Yelp 上 HR@10 从 0.1719 提至 0.1826，NDCG@10 从 0.1061 提至 0.1153，表明跨视图对齐有效改善了融合表示质量。

5) 模块间协同互补：性能提升源于多模块有机协同——语义模块增强理解，知识图谱补充结构关系，跨视图对齐缓解表示偏移，强化学习实现长期价值排序。多层次协同建模保证了模型的稳定最优效果。

3.2.3. 超参数敏感性分析

本文进一步对若干关键参数进行敏感性实验，包括折扣因子 γ、候选集规模 Top-A、知识传播层数 L

以及跨视图对齐损失权重 λ_c。实验主要报告 HR@10 与 NDCG@10 两项指标。实验结果如表 7~10 所示。

Table 7. The impact of discount factor settings on model performance
表 7. 折扣因子设置对模型性能影响

	Amazon-Book		Yelp	
γ	**HR@10**	**NDCG@10**	**HR@10**	**NDCG@10**
0.5	0.1894	0.1175	0.1732	0.1071
0.7	**0.1987**	**0.1243**	**0.1826**	**0.1153**
0.8	0.1961	0.1224	0.1808	0.1136
0.9	0.1928	0.1199	0.1779	0.1108

Table 8. The impact of candidate set size setting on model performance
表 8. 候选集规模设置对模型性能影响

	Amazon-Book		Yelp	
Top-A	**HR@10**	**NDCG@10**	**HR@10**	**NDCG@10**
200	0.1848	0.1141	0.1681	0.1037
500	0.1926	0.1205	0.1760	0.1099
1000	**0.1987**	**0.1243**	**0.1826**	**0.1153**
2000	0.1979	0.1236	0.1818	0.1147

Table 9. The influence of knowledge dissemination layer settings on model performance
表 9. 知识传播层数设置对模型性能影响

	Amazon-Book		Yelp	
L	**HR@10**	**NDCG@10**	**HR@10**	**NDCG@10**
1	0.1919	0.1198	0.1754	0.1093
2	**0.1987**	**0.1243**	**0.1826**	**0.1153**
3	0.1956	0.1217	0.1791	0.1124

Table 10. Comparing the impact of loss weight settings on model performance
表 10. 对比损失权重设置对模型性能影响

	Amazon-Book		Yelp	
λ_c	**HR@10**	**NDCG@10**	**HR@10**	**NDCG@10**
0.01	0.1931	0.1187	0.1771	0.1101
0.05	0.1968	0.1221	0.1803	0.1130
0.1	**0.1987**	**0.1243**	**0.1826**	**0.1153**
0.5	0.1942	0.1202	0.1788	0.1117
1.0	0.1885	0.1154	0.1734	0.1076

综上分析，折扣因子 γ：当 γ = 0.7 时模型最优(Amazon-Book 上 HR@10 = 0.1987)，γ 过小会忽视长期收益，过大则引入估计噪声。候选集规模 Top-A = 1000 时性能最佳，过小导致真实项目被遗漏，过大则引入噪声且增加计算开销。知识传播层数 L = 2 时最优，层数过深易导致过平滑问题。跨视图对齐损失权重 λ_c = 0.1 时最优，过小则对齐约束不足，过大则削弱推荐任务优化。

3.2.4. 收敛性分析

如图 3 所示，随着训练轮次增加，模型的训练损失在整体上平稳下降，验证集上的 HR@10 和 NDCG@10 也在不断提升，大约在 60 轮之后慢慢趋于收敛。这个趋势说明，本文模型在联合训练的过程中，优化得比较稳定。在训练刚开始的时候，模型能比较快地学到用户行为序列里的基本模式，所以损失下降和指标提升都比较明显。到了训练后期，指标提升的速度慢慢变缓，说明模型已经进入了稳定的收敛阶段。因为本文用了经验回放机制、目标网络软更新，还有 Double Dueling Q 网络结构，整个训练过程里没有出现明显的性能波动，这说明我们设计的强化学习优化框架，能有效缓解传统 Q 学习中常见的高方差和不稳定问题。另外，NDCG@10 持续增长也说明，模型不光让真正匹配的项目更容易被选入推荐列表，还在不断优化它们在列表里的排序位置。

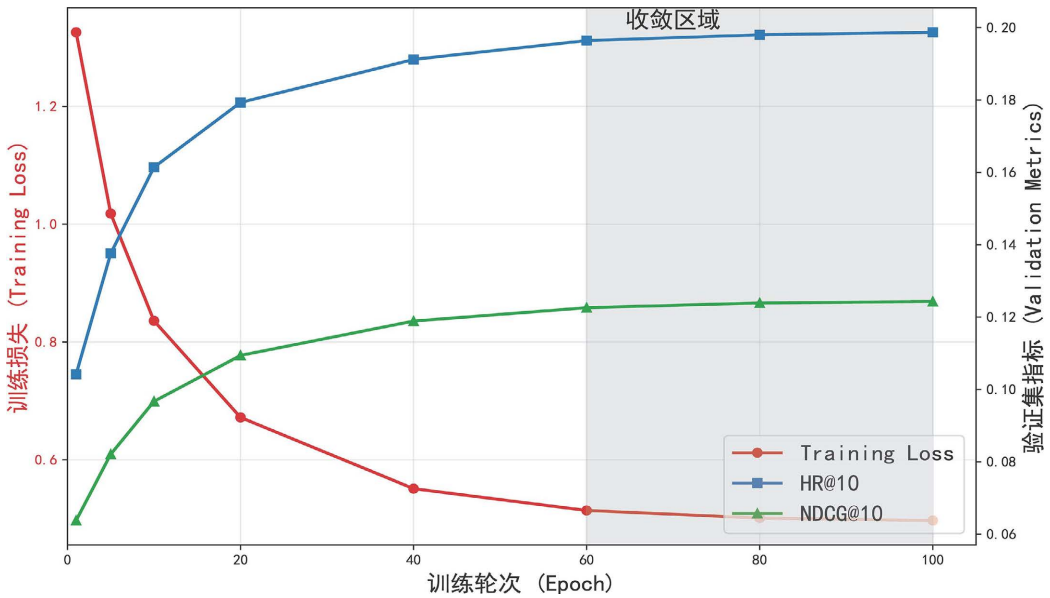

Figure 3. Visualization of the model training process
图 3. 模型训练可视化

4. 结论

本文将推荐过程建模为序列决策问题，提出融合大语言模型语义增强、知识图谱结构建模与深度强化学习的个性化推荐框架。在 Amazon-Book 和 Yelp 数据集上的实验表明，该方法在 HR@K 和 NDCG@K 等指标上均优于多种基线模型。消融实验验证了强化学习决策、语义增强、知识图谱传播及跨视图对齐四个模块的有效性。

尽管所提方法取得了较好的效果，但其在深度强化学习框架设计及应用层面仍存在若干局限性，值得在未来工作中进一步探索与改进。

首先，在当前 DRL 框架中，奖励函数设计较为简单，主要基于是否命中目标项目进行二元反馈。这种方式难以全面反映用户对推荐结果的多维满意度。未来可以考虑构建更稠密、更具表达力的奖励函数，

例如结合用户对推荐项目的显式评分、隐式反馈(如点击、停留时长、播放完成度等)，以更精准地刻画用户的真实偏好变化。

其次，当前的状态表示主要依赖于用户行为序列与语义画像，尚未显式引入上下文信息，如时间、地点、设备类型、社交情境等。尤其在 Yelp 这类生活服务场景中，用户兴趣受情境因素影响显著。未来研究可探索将时间感知编码、地理偏好嵌入等上下文特征显式地融入状态表示中，从而提升模型在复杂场景下的适应能力。

再次，本文方法对冷启动场景的适应性仍有不足。新项目缺乏历史交互与知识图谱关联，难以获得稳定的动作表示。未来可引入元学习策略，使模型能够从少量交互中快速估计新项目的嵌入；或设计结合探索–利用机制的策略，如基于不确定性估计的探索方法，在推荐过程中主动试探新项目，平衡短期收益与长期学习效率。综上所述，本文提出的融合大语言模型与深度强化学习的推荐框架在多个公开数据集上表现优异，验证了多源信息融合与序列决策建模的有效性。未来将在上述方向继续深入，推动该方法向更复杂、更动态的真实推荐场景落地。

参考文献

[1] Lyu, H., Jiang, S., Zeng, H., Xia, Y., Wang, Q., Zhang, S., *et al*. (2024) LLM-Rec: Personalized Recommendation via Prompting Large Language Models. *Findings of the Association for Computational Linguistics*: *NAACL* 2024, Mexico City, 16-21 June 2024, 583-612. https://doi.org/10.18653/v1/2024.findings-naacl.39

[2] Chung, J., Gulcehre, C., Cho, K. and Bengio, Y. (2014) Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. arXiv:1412.3555.

[3] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L. and Polosukhin, I. (2017) Attention Is All You Need. *Conference and Workshop on Neural Information Processing Systems*, Long Beach, 4-9 December 2017, 5998-6008.

[4] Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P. and Bengio, Y. (2018) Graph Attention Networks. *Proceedings of the* 6*th International Conference on Learning Representations*, Vancouver.

[5] Wang, X., He, X., Cao, Y., Liu, M. and Chua, T. (2019) KGAT: Knowledge Graph Attention Network for Recommendation. *Proceedings of the* 25*th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Anchorage, 4-8 August 2019, 950-958. https://doi.org/10.1145/3292500.3330989

[6] Chen, T., Kornblith, S., Norouzi, M. and Hinton, G. (2020) A Simple Framework for Contrastive Learning of Visual Representations. In: Daumé, P. and Singh, A., Eds., *Proceedings of the 37th International Conference on Machine Learning*, JMLR.org, 1597-1607.

[7] He, K., Fan, H., Wu, Y., Xie, S. and Girshick, R. (2020) Momentum Contrast for Unsupervised Visual Representation Learning. 2020 *IEEE*/*CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Seattle, 13-19 June 2020, 9729-9738. https://doi.org/10.1109/cvpr42600.2020.00975

[8] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., *et al*. (2015) Human-Level Control through Deep Reinforcement Learning. *Nature*, **518**, 529-533. https://doi.org/10.1038/nature14236

[9] Wang, Z., Schaul, T., Hessel, M., Van Hasselt, H., Lanctot, M. and De Freitas, F. (2016) Dueling Network Architectures for Deep Reinforcement Learning. *Proceedings of the* 33*rd International Conference on Machine Learning*, New York, 19-24 June 2016, 1995-2003.

[10] Van Hasselt, H., Guez, A. and Silver, D. (2016) Deep Reinforcement Learning with Double Q-Learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, **30**, 2094-2100. https://doi.org/10.1609/aaai.v30i1.10295