

论文《Recommender Systems with Generative Retrieval》的复现报告

摘要

现代推荐系统的普遍做法是通过模型建模学习用户和物品的表征向量，将用户和候选物品的表征向量映射到同一高维空间中。利用高效的检索算法，如最近邻搜索选择最佳的推荐物品。本报告复现的 TIGER 模型提出了一种生成式推荐方式，打破了传统的 query-search 推荐范式。具体地，该模型抛弃了基于随机 ID 的物品表示方式，使用物品文本构建物品的语义表示。采用预训练好的大语言模型作为直接推荐器，使用用户交互数据对其进行全参数微调，采用自回归的方式直接生成物品。本报告根据其论文报告的实现细节进行复现，使用了两个数据集验证该模型的推荐性能。本复现报告表明，TIGER 模型与论文中报告的性能基本一致，但性能严重依赖于文本质量；同时通过消融实验证明，传统模型中学习到的协同信号对于生成式模型同样有着至关重要的作用。

关键词：生成式序列推荐；物品文本；语义物品表示；大语言模型；协同信号；

1 引言

自 2022 年底 OpenAI 发布了 GPT3.5 [13]，各个科研领域都掀起了一股将大语言模型应用到各自领域的热潮，这其中也包括推荐系统。在传统的推荐模型中，一般使用离散无重复的 ID 来唯一表示物品。例如对于物品集 V 而言，其中的每一个物品使用 $\{1, 2, \dots, |V|\}$ 中的一个元素进行标识。并且每个 ID 有唯一的向量与之应对，通过模型的学习，建立向量与向量之间的协同关系。但在大语言模型中，其所包含的数字含义与传统推荐模型具有一定的差异性。若使用 ID 来对物品进行编码，无法有效地对物品进行表征。因此有些研究尝试使用文本当作物品的唯一标识，如 title 和 description 等字段 [3, 6]。但是对于序列推荐而言，有些用户的交互序列很长。如果使用文本对序列进行表征，会导致模型的计算时间复杂度大大增加。因此，如何在大模型中有效地对物品进行唯一标识，目前是一个很重要的研究方向。

本报告复现的 TIGER 模型 [16] 提出了使用文本构建具有语义信息的物品 ID，保证在语义上越相似的物品拥有越长的相同 ID 前缀子串。这样能够保证在计算自注意力分数时，相似物品之间具有较高的分数。同时该模型摒弃了传统的两阶段推荐范式，提出通过自回归的方式端到端地生成推荐物品。该论文由 Google 团队发表于会议 NeurIPS2023 上，属于生成式推荐的开山之作。因此验证该种新型推荐范式的有效性，有助于展开后续的相关研究。

2 相关工作

2.1 传统序列推荐模型

由于用户的行为天然具有时序性以及长短期兴趣爱好变化，研究人员发现利用用户的时序交互信息进行推荐能够产生更好的推荐结果，因此序列推荐孕育而生并逐渐成为推荐系统领域中一个热门研究方向。

在传统机器学习时代，研究主要集中于使用马尔科夫链这种可进行状态转移的模型对序列信息进行建模 [17]。随着深度学习技术的发展，RNN [4]、CNN [18]、GNN [19] 以及 Transformers [5] 等深度模型都被应用到序列推荐中。这类模型通常是使用不同的模型对用户历史行为进行建模，从不同的角度挖掘用户的潜在兴趣，以此更好地对用户和物品进行表征。此外为了解决冷启动等问题，还有些研究尝试将物品的辅助信息 [9]，多行为信息 [2, 21] 以及其他域 [22] 的信息考虑到建模过程中，以达到更好的推荐性能。

2.2 基于大语言模型的推荐模型

基于大语言模型的推荐模型按照在推荐系统中的主次作用，可以分成两类：通用推荐器和特征编码器。并且这两类使用方法都可以根据是否进行微调而进一步划分。

对于通用推荐器而言，大语言模型直接产生推荐结果。早期的研究主要关注于将推荐任务转为自然语言的形式，让大模型生成用户可能感兴趣的物品 [7, 10]。结果表明，不进行调整的方式虽然能够优于随机推荐，但性能不如传统的 ID 推荐范式 [10]。其主要原因是预训练好的大语言模型无法理解推荐中的协同信号，因此有些研究尝试通过微调让模型理解协同信号从而进一步提升性能 [1, 8]。由于大语言模型无法直接生成确切且真实存在的物品，因此在这些方法中往往需要在提示词中加入一个由传统模型生成的物品列表。基于此，上述的研究并不完全属于生成式推荐范式的范畴。

对于特征编码器而言，大语言模型充当传统推荐模型的一个组件。非微调的做法通常是利用预训练好的大模型将文本或图像转换成向量，并将这些向量作为传统推荐模型中对应物品的表征向量 [14]。此外随着生成式大模型的出现，也有研究尝试利用大模型的强大推理能力以及丰富的现实世界知识生成与物品相关的文本内容作为辅助信息 [20]。而微调的方法则是一种端到端的建模方法，将特征编码器与传统推荐模型拼接起来。在训练过程更新语言模型的参数，使二者在推荐任务中协同工作 [23]。虽然这种方式会取得更好的效果，但是需要耗费大量的计算资源 [23]。

3 本文方法

3.1 本文方法概述

TIGER 模型包含三个步骤：文本编码，表征离散化以及有监督微调。对于整个模型而言，三个阶段为独立、按顺序进行的。在下列章节中，将详细介绍相关内容。

3.2 文本编码

由于无法将文字直接作为模型的输入，因此第一步要先将文本转换为向量。在这个过程中需要使用的预训练好的文本编码器对文本进行编码。

给定一个物品 i 的相应文本内容 $C_i = \{w_1, w_2, \dots, w_m\}$ ，其中 w_i 表示第 i 个 token。将 C_i 输入到文本编码器中，每个位置的 token 都将得到一个输出向量。因此所有 token 的输出构成该物品的文本表示矩阵 $\mathbf{E} \in R^{m \times d}$ ，其中 d 是文本编码器预定义好的输出维度。对该矩阵的行做平均池化，得到单个物品的语义向量 $\mathbf{x} \in R^{1 \times d}$ 。

$$\mathbf{E} = \text{TextEncoder}(C_i) \quad (1)$$

$$\mathbf{x} = \text{Mean}(\mathbf{E}) \quad (2)$$

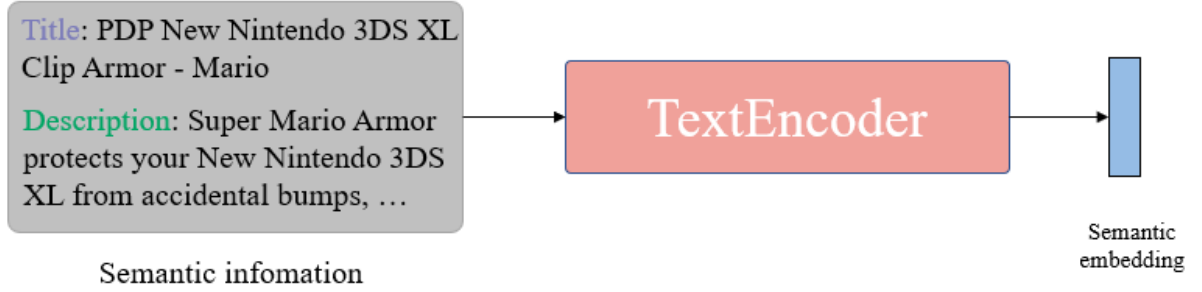


图 1. 文本编码示意图

3.3 表征离散化

生成式推荐的主要思想在于，在语义上越相似的物品在表征为离散化 ID 时具有越多的相同 ID 前缀。由于大模型的本质仍然是自注意力机制，当相同 ID 前缀更多时，相似物品与相似物品计算得到的注意力分数就会更高。这与相似物品会被相似用户交互的协同信号理念一致。因此为了达成这个目的，在原文中引入了 RQ-VAE 的技术。

虽然该技术中包含了 VAE 的字段，但其本质是 Auto Encoder。首先在 3.2 中获得的语义向量 \mathbf{x} 会经过由多层全连接层组成的 *Encoder*，从一个高维向量转换为低维的隐语义向量 $\mathbf{z} \in R^{1 \times d'}$ ，其中 d' 是最后一层全连接层的输出维度。

$$\mathbf{z} = \text{Encoder}(\mathbf{x}) \quad (3)$$

此外，RQ-VAE 中设计了 D 层的码本，每个码本的大小为 K ，第 d 层的码本可以表示为 $C_d = \{\mathbf{e}_i \mid \mathbf{e}_i \in R^{d'}\}_{i=1}^K$ 。在每一层中会使用中间变量 \mathbf{r}_l 去对应的码本中寻找距离最小的向量。索引得到的向量记为 \mathbf{e}_{c_d} ，其中 c_d 为该向量在码本中的位置。在进入下一层之前，利用公式 5 更新中间变量。通过残差操作，逐步逼近 \mathbf{z} 。特别的， $\mathbf{r}_0 = \mathbf{z}$ 。

$$c_d = \text{argmin}_i \|\mathbf{r}_d - \mathbf{e}_i\|^2 \quad (4)$$

$$\mathbf{r}_{d+1} = \mathbf{r}_d - \mathbf{e}_{c_d} \quad (5)$$

经过 D 层的哈希索引后，将每层索引得到的向量相加得到复原的隐向量 $\hat{z} \in R^{1 \times d'}$ 。通过 Decoder，得到复原的物品语义向量 $\hat{x} \in R^{1 \times d}$ 。这里 Decoder 即为 Encoder 的逆序全连接层。

$$\hat{z} = \sum_{d=1}^D e_{cd} \quad (6)$$

$$\hat{x} = \text{Decoder}(\hat{z}) \quad (7)$$

模型的损失函数如公式 (8) 所示。其中前一项为常规的自编码器的重构损失函数。后一项为 RQ-VAE 中引入的损失，其本质式希望让每一次去索引的向量与索引得到的向量越接近越好。但是由于索引这个操作在深度学习中是不可微的，导致无法利用反向传播对码本进行训练。因此引入了 sg 操作（梯度停止反向传播，stop gradient），通过近似的方法使得任意操作可微。整体的模型结构如图 2 所示。

$$\mathcal{L} = \|x - \hat{x}\|^2 + \sum_{d=1}^D \{ \|sg[r_d] - e_{cd}\|^2 + \beta \|r_d - sg[e_{cd}]\|^2 \} \quad (8)$$

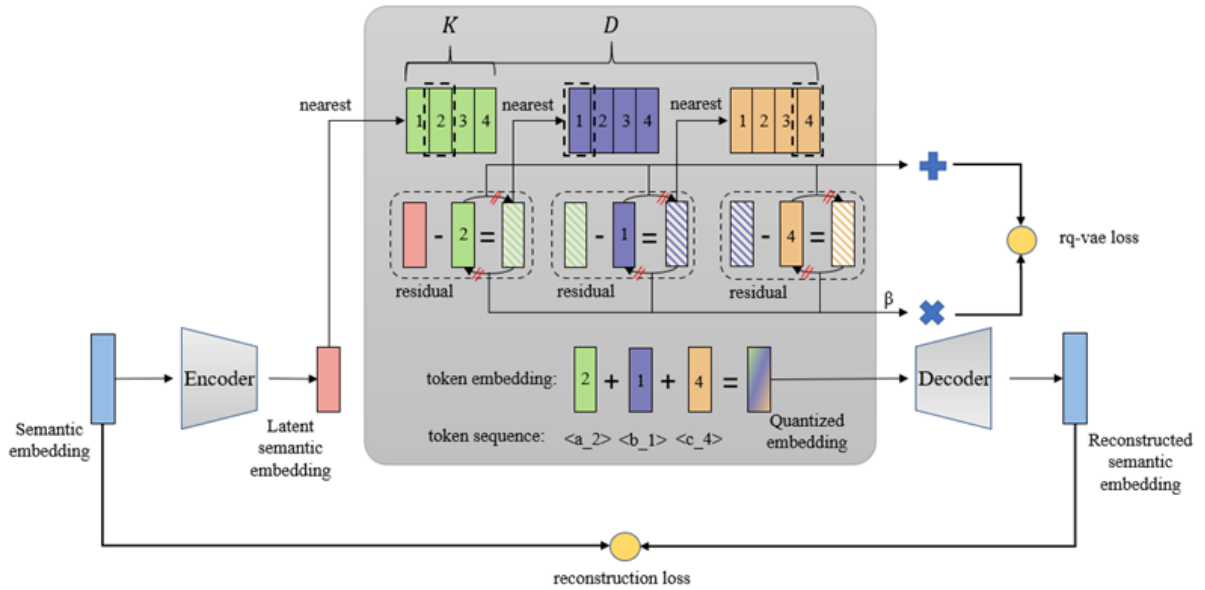


图 2. RQ-VAE 结构示意图

3.4 有监督微调

在进行上述步骤后，一个物品可以使用形如 $\langle a_{41} \rangle \langle b_{18} \rangle \langle c_4 \rangle$ 的 ID 序列进行表示。因此模型的输入，将由传统模型中长度为 $maxlen$ 的 ID 序列将转变为长度为 $maxlen \times D$ 的 ID 序列，其中 $maxlen$ 为序列长度超参数。在进行微调前，要先将 4.2 小节中得到物品离散编码全部加入到词表中。因此大模型词表变为原来的词表加上 4.2 中得到的编码集合大小。新加入编码的对应向量会被进行初始化，并在训练过程得到优化。扩充过后的词表，会将一个物品拆分为 $\langle a_{41} \rangle$ 、 $\langle b_{18} \rangle$ 和 $\langle c_4 \rangle$ 三个离散 ID。

在训练过程中，转变后的 ID 序列作为 Encoder 的输入，验证集作为 Decoder 的监督信号。训练过程使用全参数微调。微调阶段的损失如式 (9) 所示。其中 y_t 为第 t 步产生的输

出, \mathbf{x} 为 Encoder 的输入, $y_{<t}$ 为 Decoder 在时间步 t 之前产生的所有输出。微调阶段的示意图如图 3 所示。

$$\mathcal{L} = - \sum_{t=1}^{|y|} \log P_{\theta}(y_t | y_{<t}, \mathbf{x}) \quad (9)$$

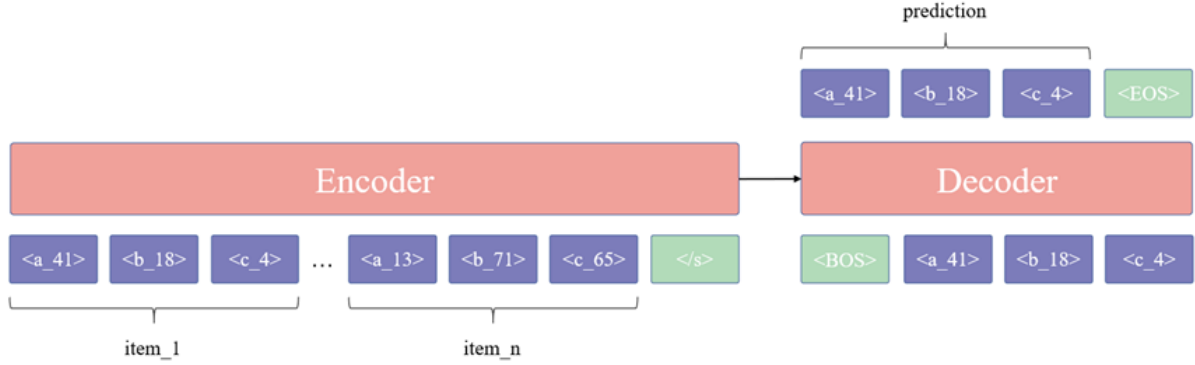


图 3. 微调阶段示意图

4 复现细节

4.1 与已有开源代码对比

本报告实现的 TIGER 模型并没有公布任何源码, 但是本文复现代码中的 RQ-VAE 部分的代码来源于另外一篇生成式序列推荐论文 LC-Rec [26], 其余实现细节均为自己复现。在本次复现中没有对原模型进行创新, 主要是对该种新型推荐范式性能的验证。本报告的复现围绕以下研究问题, 做了一系列实验进行报告:

研究问题 1: TIGER 模型能否达到其论文报告的结果?

研究问题 2: 生成式推荐范式与传统推荐范式性能相比如何?

研究问题 3: 不同的物品表征方式对生成式推荐范式有何种影响?

4.2 数据集和评估方式

在本报告中, 采用了两个开源数据集 Bili_Cartoon [24] 和 Amazon Games 数据集 [12]。其中 Bili_Cartoon 数据集包含视频的中英文标题, Amazon Games 数据集包含了若干文本信息 (如标题, 描述, 种类等)。数据集的统计特征如表 1 所示。

采用了 $HR@5, 10$ 和 $NDCG@5, 10$ 共四个指标在测试集上进行验证。由于生成式推荐范式推理速度慢, 无法在验证集上采用上述评估指标。因此使用了在验证集上的 loss 值进行判断, 保存在验证集上的 loss 最低的模型。并且若 loss 超过 10 个 epoch 没有下降, 则停止训练。采用了留一法划分验证集和测试集。

表 1. 数据集统计特征

数据集	用户数量	物品数量	交互数量	稀疏度	平均序列长度
Bili_Cartoon	30300	4724	215443	99.85	7.11
Amazon Games	50547	16860	452989	99.95	8.96

4.3 模型实现细节

在文本编码阶段，本文的文本编码器采用了 SentenceT5 [11] 作为文本编码器，其输出向量维度 $d = 768$ 。

在表征离散化阶段，码本数量 $D = 3$ ，一个码本中的向量数 $K = 256$ 。学习率设置为 0.001，迭代训练 5000 次并且数据批次大小为 512，使用了 AdamW 优化器。使用 constant 学习率调度器，并将 warmup 轮次设置为 50。损失函数中的 $\beta = 0.25$ ，Encoder 中每个全连接层的输出大小为 [2048, 1024, 512, 256, 128, 64]，即 $d' = 64$ 。需要注意的是，为了防止在训练初期出现大部分隐语义向量哈希到同一个向量上的现象，在使用第一个批次数据训练时使用聚类对每一层码本的向量进行初始化。具体来说，将向量聚成 K 个类，每个类的均值向量初始化为其中一个码本向量。

在微调阶段，使用了 T5-base [15] 作为推荐器。学习率和正则化系数分别设置为 $5e-4$ 和 0.01，迭代训练 200 次并且数据批次大小为 128，使用了 AdamW 优化器。使用 cosine 学习率调度器，并将 *warmup_ratio* 设置为 0.01。交互序列的最大长度设置为 20。

推理时，将超参数 *num_beams* 设置为 50，并设置了字典树限制模型的输出。

5 实验结果分析

5.1 复现结果

由于无法获取 TIGER 原文 [16] 所使用的数据集，为了让实验具有可比性，本次复现中的 Amazon Games 数据采用了论文 [26] 中提供的数据集，并且与该论文中报告的 TIGER 性能进行对比。从表 2 的复现结果可以看到，复现效果与其报告的性能基本一致，证明了生成式推荐范式的可行性。

表 2. 复现结果

数据集	模型	HIT@5	HIT@10	NDCG@5	NDCG@10
Amazon Games	LC-Rec	0.0599	0.0939	0.0392	0.0501
	自己复现	0.0606	0.0987	0.0398	0.0519

5.2 模型性能比较

除了验证该范式的可行性，本文还将新旧两种推荐范式进行了性能上的对比。需要注意的是，Bili_Cartoon 数据集为我自己使用的数据集，在原文中没有使用。此外，表 2 中的报

告的 TIGER 并非最佳效果，而是按照论文细节进行复现得到的效果。在本实验中，选取了序列推荐中最为出名的 SASRec 模型 [5]，该模型为推荐领域共同认可的模型。表 2 中所有对比模型结果均由伯乐框架获得 [25]。

表 3. 模型性能比较

数据集	模型	HIT@5	HIT@10	NDCG@5	NDCG@10
Bili_Cartoon	SASRec	0.0802	0.1318	0.0475	0.0626
	TIGER	0.0529	0.0767	0.0376	0.0453
Amazon Games	SASRec	0.0616	0.1011	0.0353	0.0480
	TIGER	0.0606	0.0987	0.0398	0.0519

通过表 2 可以观察到，在 Bili_Cartoon 数据集上生成式推荐范式得到的效果均比 ID 推荐范式差，并且差距在 26.3%(NDCG@5)-71.8%(HIT@10)。而在 Amazon Games 数据集上，两个范式的效果基本一致。并且在 NDCG 这个指标上，生成式推荐范式效果更好。通过深入观察数据，可以得到出现这种现象的原因：1. Bili_Cartoon 数据集上的文本质量较差，有些视频的标题并不能很好地概括视频的内容，导致模型无法简单地从视频标题学习到用户的潜在兴趣。2. 对于 Bili_Cartoon 在生成语义向量的时候，只使用了标题字段。而 Amazon Games 使用的文本更长，生成了质量更好的语义向量。通过这个简单的对比实验，我们可以发现生成式推荐范式的推荐性能与文本的质量相关。

5.3 消融实验

为了验证不同方式的物品表征对生成式推荐的性能有什么影响，在消融实验中设计了五种变体。其中带有 pretrained 字段的模型，在 RQ-VAE 阶段将物品的语义向量替换为 SASRec 中训练后的物品向量。带有 nc 字段的模型，解决了哈希冲突问题。具体来说，由于经过离散化的物品可能存在哈希冲突，即对于不同的物品均映射为同一个离散 ID，添加额外一个 ID 解决冲突问题。例如对于原始 ID 为 1 和 2 的物品均映射为 $\langle a_{41} \rangle \langle b_{18} \rangle \langle c_4 \rangle$ ，那么物品 1 将表示为 $\langle a_{41} \rangle \langle b_{18} \rangle \langle c_4 \rangle \langle d_0 \rangle$ ，而物品 2 将表示为 $\langle a_{41} \rangle \langle b_{18} \rangle \langle c_4 \rangle \langle d_1 \rangle$ 。而对于那些不存在冲突的物品，同样添加额外的 ID 来保证所有物品的 ID 长度一致。此外考虑到模型的训练开销较大，消融实验只在 Bili_Cartoon 数据集上进行。

表 4. 消融实验

变体名称	HIT@5	HIT@10	NDCG@5	NDCG@10
TIGER-pretrained	0.0707	0.1081	0.0469	0.0589
TIGER-SID	0.0529	0.0767	0.0376	0.0453
TIGER-pretrained-nc	0.0715	0.1076	0.0473	0.0589
TIGER-SID-nc	0.0559	0.0854	0.0346	0.0442
TIGER-SID(K=256, D=4)	0.0552	0.0901	0.0362	0.0474

通过观察表 3 可以发现,对于变体 TIGER-pretrained(-nc) 和 TIGER-SID(-nc) 而言,使用预训练得到的物品表示能够有效提升推荐性能。这个现象表明仅仅使用文本上的相似度来近似协同信号是远远不够的,特别是当文本质量不好时。对于变体 TIGER-pretrained(SID)/TIGER-pretrained(SID)-nc 而言,是否存在哈希冲突对模型的性能影响并不大。对于变体 TIGER-SID 和 TIGER-SID(K=256, D=4) 而言,可以观察到适当增加 RQ-VAE 的残差层数有助于推荐性能提升。

6 总结与展望

本报告复现了一种生成式推荐范式的序列推荐方法-TIGER 模型。本报告根据其论文细节并参考了部分他人代码复现模型,其复现结果在 Amazon Games 数据集上接近论文 [26] 报告的结果。该现象表明,生成式推荐范式有一定的可行性。

同时本报告还使用了另外的开源数据集验证其性能,得到如下结论:当物品文本质量较低时,生成式推荐的效果会大打折扣,这也是大多数单纯使用文本进行推荐的通病。此外在训练以及推理过程中,时间和资源开销也会更大。本文还做了充分的消融实验,设计了不同的物品表征方式研究其对模型性能的影响。得到了如下的结论:ID 范式依然具有较大的统治力,利用其生成的物品表征有助于提高模型的性能提升。

在后续的研究中,可以考虑在两个方向上进一步改进模型:(1) 如何使用 ID 范式助力生成范式提高推荐性能,特别是当文本质量较低时,引入 ID 范式是非常重要的。(2) 如何加速模型的训练和推理速度。

参考文献

- [1] Keqin Bao, Jizhi Zhang, Yang Zhang, Wenjie Wang, Fuli Feng, and Xiangnan He. Tallrec: An effective and efficient tuning framework to align large language model with recommendation. In *Proceedings of the 17th ACM Conference on Recommender Systems*, pages 1007–1014, 2023.
- [2] Ming Chen, Weike Pan, and Zhong Ming. Explicit and implicit modeling via dual-path transformer for behavior set-informed sequential recommendation. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD '24*, page 329–340, New York, NY, USA, 2024. Association for Computing Machinery.
- [3] Shijie Geng, Shuchang Liu, Zuohui Fu, Yingqiang Ge, and Yongfeng Zhang. Recommendation as language processing (rlp): A unified pretrain, personalized prompt & predict paradigm (p5). In *Proceedings of the 16th ACM Conference on Recommender Systems, RecSys '22*, page 299–315, New York, NY, USA, 2022. Association for Computing Machinery.
- [4] Balázs Hidasi and Alexandros Karatzoglou. Recurrent neural networks with top-k gains for session-based recommendations. In *Proceedings of the 27th ACM International Conference*

- on *Information and Knowledge Management*, CIKM '18, page 843–852, New York, NY, USA, 2018. Association for Computing Machinery.
- [5] Wang-Cheng Kang and Julian McAuley. Self-attentive sequential recommendation. In *2018 IEEE International Conference on Data Mining (ICDM)*, pages 197–206, 2018.
 - [6] Jiacheng Li, Ming Wang, Jin Li, Jinmiao Fu, Xin Shen, Jingbo Shang, and Julian McAuley. Text is all you need: Learning language representations for sequential recommendation. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '23, page 1258–1267, New York, NY, USA, 2023. Association for Computing Machinery.
 - [7] Jinming Li, Wentao Zhang, Tian Wang, Guanglei Xiong, Alan Lu, and Gerard Medioni. Gpt4rec: A generative framework for personalized recommendation and user interests interpretation. *arXiv preprint arXiv:2304.03879*, 2023.
 - [8] Jiayi Liao, Sihang Li, Zhengyi Yang, Jiancan Wu, Yancheng Yuan, Xiang Wang, and Xiangnan He. Llara: Aligning large language models with sequential recommenders. *arXiv preprint arXiv:2312.02445*, 2023.
 - [9] Xiaolin Lin, Jinwei Luo, Junwei Pan, Weike Pan, Zhong Ming, Xun Liu, Shudong Huang, and Jie Jiang. Multi-sequence attentive user representation learning for side-information integrated sequential recommendation. In *Proceedings of the 17th ACM International Conference on Web Search and Data Mining*, WSDM '24, page 414–423, New York, NY, USA, 2024. Association for Computing Machinery.
 - [10] Junling Liu, Chao Liu, Peilin Zhou, Renjie Lv, Kang Zhou, and Yan Zhang. Is chatgpt a good recommender? a preliminary study. *arXiv preprint arXiv:2304.10149*, 2023.
 - [11] Jianmo Ni, Gustavo Hernandez Abrego, Noah Constant, Ji Ma, Keith B Hall, Daniel Cer, and Yinfei Yang. Sentence-t5: Scalable sentence encoders from pre-trained text-to-text models. *arXiv preprint arXiv:2108.08877*, 2021.
 - [12] Jianmo Ni, Jiacheng Li, and Julian McAuley. Justifying recommendations using distantly-labeled reviews and fine-grained aspects. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)*, pages 188–197, 2019.
 - [13] Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. Training language models to follow instructions with human feedback. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, NIPS '22, Red Hook, NY, USA, 2024. Curran Associates Inc.

- [14] Zhaopeng Qiu, Xian Wu, Jingyue Gao, and Wei Fan. U-bert: Pre-training user representations for improved recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 4320–4327, 2021.
- [15] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. Exploring the limits of transfer learning with a unified text-to-text transformer. *J. Mach. Learn. Res.*, 21(1), January 2020.
- [16] Shashank Rajput, Nikhil Mehta, Anima Singh, Raghunandan Keshavan, Trung Vu, Lukasz Heidt, Lichan Hong, Yi Tay, Vinh Q. Tran, Jonah Samost, Maciej Kula, Ed H. Chi, and Maheswaran Sathiamoorthy. Recommender systems with generative retrieval. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, NIPS ’23, Red Hook, NY, USA, 2024. Curran Associates Inc.
- [17] Steffen Rendle, Christoph Freudenthaler, and Lars Schmidt-Thieme. Factorizing personalized markov chains for next-basket recommendation. In *Proceedings of the 19th International Conference on World Wide Web*, WWW ’10, page 811–820, New York, NY, USA, 2010. Association for Computing Machinery.
- [18] Jiayi Tang and Ke Wang. Personalized top-n sequential recommendation via convolutional sequence embedding. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, WSDM ’18, page 565–573, New York, NY, USA, 2018. Association for Computing Machinery.
- [19] Shu Wu, Yuyuan Tang, Yanqiao Zhu, Liang Wang, Xing Xie, and Tieniu Tan. Session-based recommendation with graph neural networks. In *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*, AAAI’19/IAAI’19/EAAI’19. AAAI Press, 2019.
- [20] Yunjia Xi, Weiwen Liu, Jianghao Lin, Xiaoling Cai, Hong Zhu, Jieming Zhu, Bo Chen, Ruiming Tang, Weinan Zhang, and Yong Yu. Towards open-world recommendation with knowledge augmentation from large language models. In *Proceedings of the 18th ACM Conference on Recommender Systems*, pages 12–22, 2024.
- [21] Jing Xiao, Weike Pan, and Zhong Ming. A generic behavior-aware data augmentation framework for sequential recommendation. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR ’24, page 1578–1588, New York, NY, USA, 2024. Association for Computing Machinery.
- [22] Zitao Xu, Weike Pan, and Zhong Ming. A multi-view graph contrastive learning framework for cross-domain sequential recommendation. In *Proceedings of the 17th ACM Conference on Recommender Systems*, RecSys ’23, page 491–501, New York, NY, USA, 2023. Association for Computing Machinery.

- [23] Zheng Yuan, Fajie Yuan, Yu Song, Youhua Li, Junchen Fu, Fei Yang, Yunzhu Pan, and Yongxin Ni. Where to go next for recommender systems? id-vs. modality-based recommender models revisited. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2639–2649, 2023.
- [24] Jiaqi Zhang, Yu Cheng, Yongxin Ni, Yunzhu Pan, Zheng Yuan, Junchen Fu, Youhua Li, Jie Wang, and Fajie Yuan. Ninerec: A benchmark dataset suite for evaluating transferable recommendation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.
- [25] Wayne Xin Zhao, Shanlei Mu, Yupeng Hou, Zihan Lin, Yushuo Chen, Xingyu Pan, Kaiyuan Li, Yujie Lu, Hui Wang, Changxin Tian, Yingqian Min, Zhichao Feng, Xinyan Fan, Xu Chen, Pengfei Wang, Wendi Ji, Yaliang Li, Xiaoling Wang, and Ji-Rong Wen. Recbole: Towards a unified, comprehensive and efficient framework for recommendation algorithms. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management, CIKM '21*, page 4653–4664, New York, NY, USA, 2021. Association for Computing Machinery.
- [26] Bowen Zheng, Yupeng Hou, Hongyu Lu, Yu Chen, Wayne Xin Zhao, Ming Chen, and Ji-Rong Wen. Adapting large language models by integrating collaborative semantics for recommendation. In *2024 IEEE 40th International Conference on Data Engineering (ICDE)*, pages 1435–1448. IEEE, 2024.