arXiv:2304.03879v1 [cs.IR] 8 Apr 2023

GPT4Rec：用于个性化推荐和用户兴趣解读的生成式框架

李金明[1], 张文涛[2], 王天[2], 熊光磊[2], 陆艾伦[2], Gerard Medioni[2]lijinmin@umich.edu,
{wentazha,wangtan,glx,alalu,medioni}@amazon.com[1]密歇根大学，美国安娜堡[2]亚马逊，美国

摘要

自然语言处理（NLP）的最新进展推动了基于自然语言处理的推荐系统的发展，并展现出卓越的性能。然而，当前模型通常仅将物品视为普通 ID，并采用判别式建模，因此存在以下局限：（1）未能充分利用物品的内容信息以及自然语言处理模型的语言建模能力；（2）未能解读用户兴趣以提升相关性和多样性；（3）难以适应物品库存不断增长等实际情况。为解决这些局限，我们提出 GPT4Rec，这是一种受搜索引擎启发的新颖且灵活的生成式框架。该框架首先根据用户历史记录中的商品标题生成假设性的“搜索查询”，然后通过搜索这些查询来检索推荐物品。它通过在语言空间中同时学习用户和物品嵌入，克服了以往的局限。为充分捕捉用户兴趣在不同方面和粒度上的差异，从而提升相关性和多样性，我们提出了一种结合束搜索的多查询生成技术。生成的查询自然地构成了用户兴趣的可解释表示，也可用于搜索并推荐冷启动物品。结合 GPT-2 语言模型和 BM25 搜索引擎后，我们的框架在两个公开数据集上的 Recall@K 分别比最先进的方法高出 75.7% 和 22.2%。实验进一步表明，结合束搜索的多查询生成既能提升检索物品的多样性，也能提高对用户多重兴趣的覆盖率。我们还通过定性案例研究探讨了生成查询的适应性和可解释性。

CCS 概念

• **信息系统** →**推荐系统**。

关键词

个性化推荐，用户兴趣建模，生成式语言模型，查询生成，搜索

ACM 参考格式：
李金明[1], 张文涛[2], 王天[2], 熊光磊[2], 陆艾伦[2], Gerard Medioni[2]。2018年。GPT4Rec：一种用于个性化推荐和用户兴趣解读的生成式框架。载于论文集。请确保输入权利确认邮件中的正确会议名称

允许免费复制本作品的全部或部分内容，用于个人或课堂用途，前提是不得为营利或商业利益制作或分发副本，且副本首页须保留本声明及完整引文。对于本作品中由 ACM 以外方所有的组成部分，必须尊重其版权。注明出处后，允许摘录。其他复制、转载、发布到服务器或分发至列表，均须事先获得具体许可和/或支付费用。请向 permissions@acm.org 申请许可。会议缩写“*XX*”，*2018*年*6*月*3*日至*5*日，纽约州伍德斯托克© 2018年计算机协会。ACM ISBN 978-1-4503-XXXX-X/18/06...$15.00 https://doi.org/XXXXXXX.XXXXXXX

（会议缩写“*XX*”）。ACM，美国纽约州纽约市，共5页。https://doi.org/XXXXXXX.XXXXXXX

1 引言

推荐系统是信息过滤系统，可为特定用户提供个性化的相关商品建议，广泛应用于电子商务和社交媒体服务等场景 [4, 11, 14, 19]。随着自然语言处理（NLP）的发展，许多基于 NLP 的模型通过建模用户与商品的交互序列来实现个性化推荐 [9, 27, 30]。近期，大型语言模型（LLM， [1, 5, 17, 24]）在各种自然语言处理任务和对话任务（如 ChatGPT [16]）中取得了卓越表现，推动了基于 LLM 的推荐系统研究 [6, 10, 15, 22, 23, 29]。具体而言，BERT4Rec [23] 采用了 BERT [5] 的双向自注意力结构，其表现优于其他基于 NLP 的模型和序列模型 [9, 10, 18]。

尽管模型性能有所提升，现有基于自然语言处理的模型 [6, 10, 15, 22, 23, 29] 通常将物品仅视为 ID，并采用判别式建模，存在以下局限：首先，这些模型无法充分利用物品的内容信息以及自然语言处理模型的语言建模能力。其次，这些模型无法适应不断变化和增长的物品库存；这两者都是重要的实际问题 [12, 21]。此外，判别式模型难以解释用户兴趣，而这对于提升推荐的多样性和质量至关重要 [2, 3, 13, 28]。

为解决这些局限，我们提出 GPT4Rec，这是一种新颖、灵活的个性化推荐生成式框架，同时能够解读用户兴趣。我们的框架受到搜索引擎的启发：GPT4Rec 首先使用生成式语言模型生成假设的“搜索查询”，并将用户历史记录中的商品标题与生成提示结合起来作为模型输入。随后，GPT4Rec 通过搜索引擎检索这些生成的查询，从而获取推荐商品。GPT4Rec 的核心组件是一个强大的生成式语言模型，它能够在语言空间中学习用户和物品的嵌入，使我们得以利用商品标题中的语义信息，并捕捉用户的多样化兴趣。为了解析用户的多重兴趣并提升推荐多样性，我们在查询生成过程中采用多查询束搜索技术。这些查询通俗易懂，可独立用于解读用户兴趣。此外，搜索代表用户兴趣的查询以进行推荐，可以自然地解决商品冷启动问题和商品库存变化问题。最后但同样重要的是，GPT4Rec 在实践中具有良好的适用性，能够灵活地结合更先进的大语言模型或搜索引擎来提升性能。

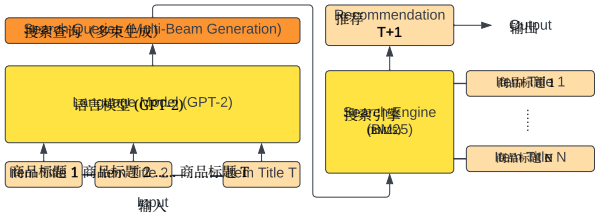


图 1：GPT4Rec 框架的架构。

总而言之，我们的贡献包括：(1) 提出 GPT4Rec，这是一种新颖、灵活的生成式框架，将推荐任务视为查询生成 + 搜索；(2) 采用多查询束搜索策略，生成多样且可解释的用户兴趣表示；(3) 在两个公开数据集上开展实验，结果表明，与最先进的方法相比，Recall@K 有了显著提升；(4) 通过定量和定性分析，我们探究并展示了生成查询的数量会提升检索物品的多样性，以及对用户多重兴趣的覆盖率。

2 方法

所提出框架的架构如图 1 所示。给定用户的商品交互序列，GPT4Rec 会通过提示词对商品标题进行格式化，并使用生成式语言模型在语言空间中学习商品和用户嵌入。随后，模型生成多个代表用户兴趣的查询，并将其输入搜索引擎，以检索用于推荐的物品。本文选用 GPT-2 语言模型和 BM25 搜索引擎，但该框架也可灵活地集成更先进的模型。

2.1 使用语言模型生成查询

GPT4Rec 的第一个组件是生成式语言模型，其目标是从商品交互序列中学习用户在语言空间中的表示，然后生成多个代表用户兴趣的查询。对所选 GPT-2 模型进行微调即可满足我们的需求；该模型拥有 1.17 亿个参数，采用复杂的 Transformer 类型架构，并在海量语言语料库上进行预训练，能够捕捉用户兴趣和商品内容信息。经过实验，我们决定使用以下提示词来格式化模型输入：

> 此前，客户购买过：<商品标题 *1*>。 <
> 商品标题 *2*>.....未来，客户想要购买

其中包含商品标题中的语义信息，并针对每位用户 u表示为 W^u。GPT-2 从 W_u 中学习用户表示，随后能够以序列方式根据条件分布 $P(\cdot|W_u)$生成查询。

为了充分表征用户的多样化兴趣并提高推荐结果的多样性，我们提出使用束搜索技术生成多个查询，而非仅生成一个查询 [18, 25]。给定束宽 m、生成评分函数 $S(\cdot)$，以及长度为 l 的候选查询 $Q^l = (q_1^l, q_2^l, ...q_m^l)$，束搜索算法会将 Q^{l+1} 更新为长度为 $l+1$、使 $\mathrm{S}(W^u, q)$ $q \in \{|q| = l+1, q_{[1:l]} \in Q^l\}$ 最大的前 m 个查询。最重要的是，这种束搜索策略能够生成在不同方面和粒度上各具多样性的用户兴趣表征。

2.2 使用搜索引擎检索商品

所提出框架的第二个组件是一个充当“判别器”的搜索引擎。它将生成的每个查询作为输入，并通过匹配评分函数，从库存中检索最相关的物品作为输出。匹配评分函数用于衡量语言空间中的相似度，在作用上类似于向量嵌入方法中的内积相似度 [4, 19]。我们选择采用 BM25 匹配评分函数[20], ，因为它是应用最广泛的基线搜索引擎之一，能够通过两个相应参数 k_1 和 b 处理词频饱和与文档长度。

令 K 表示推荐项目的总数，m 表示生成的查询数量。我们提出一种基于排序的策略，用于合并各查询的搜索结果。该策略首先从生成得分最高的查询的搜索结果中选取排名前K/m 的物品，然后按照得分排名，依次从其余查询中添加 K/m 个不重复的物品。所提出的策略能够兼顾检索物品的相关性和多样性。

2.3 训练策略

我们提出一种灵活的两步训练流程，分别优化语言模型和搜索引擎，这有利于实际中的模型迭代。给定每位用户 u的物品交互序列 i_1、i_2、.....、i_T，我们取第 2.1 节所示提示中的前 $T-1$ 个商品标题，再将其与最后一个物品 i_T 的标题拼接，构成训练语料，用于微调预训练的 GPT-2 模型。这一策略受到对比学习文献 [8], 的启发，其核心思想是：最细粒度且准确的搜索查询就是目标商品标题本身。训练完语言模型后，我们通过网格搜索参数 k_1 和 b 来优化 BM25 参数，使搜索生成的查询以检索目标物品时取得最佳性能。

3 实验

3.1 实验设置

3.1.1 数据。我们在两个公开数据集上开展实验：亚马逊美妆和电子产品类别的 5-core 亚马逊评论数据[7]。我们剔除了标题缺失或含噪声（超过 400 个字符）的商品。对每位用户的商品交互序列进行去重，并截断至最大长度 15。预处理后数据集的描述性统计信息见表2。按照下一商品预测任务的设置，[23], 将用户序列按 0.8/0.1/0.1 的比例划分为训练集、验证集和测试集，并将每个测试序列中的最后一个商品作为预测目标。

3.1.2 评估指标。 **Recall@K** 是下一商品预测中衡量兴趣的主要指标，用于衡量前K个推荐商品是否包含目标商品，并对所有用户取平均值。除了 Recall@K，我们还关注：

表 1：基线方法与所提出框架在不同生成查询数量下的总体性能。最佳性能以粗体标出，最佳基线结果以下划线标出。

数据集	召回率@K	FM-BPR	ContentRec	YouTubeDNN	BERT4Rec	GPT4Rec			
						5 个查询	10 个查询	20个查询	40个查询
美妆	K=5	0.0356	0.0254	0.0376	0.0355	**0.0653**	—	—	—
	K=10	0.0499	0.0440	0.0549	0.0513	0.0810	**0.1036**	—	—
	K=20	0.0716	0.0644	0.0753	0.0816	0.1027	0.1252	**0.1454**	—
	K=40	0.1040	0.0952	0.1066	0.1161	0.1297	0.1522	0.1743	**0.2040**
电子产品	K=5	0.0345	0.0241	0.0352	0.0362	**0.0434**	—	—	—
	K=10	0.0387	0.0307	0.0435	0.0451	0.0480	**0.0545**	—	—
	K=20	0.0441	0.0391	0.0539	0.0573	0.0524	0.0607	**0.0705**	—
	K=40	0.0505	0.0494	0.0684	0.0751	0.0574	0.0672	0.0788	**0.0918**

表2：数据集统计信息。

Name	#用户	#商品	#交互	平均长度	元信息
美妆	22,254	11,778	190,726	7.439	类别
电子产品	728,719	159,456	6,724,382	7.797	类别、品牌

多样性@K [3]，用于衡量物品之间的不相似性

$$\text{Diversity@}K = \text{Average}\left[1 - \frac{\sum_{i_1 \neq i_2} \text{Sim}(i_1, i_2)}{K(K-1)}\right],$$

其中，我们采用商品类别或品牌上的 *Jaccard* 相似度 [1] **覆盖率@K**，用于衡量推荐商品R在类别或品牌方面 U 覆盖用户序列的程度：

$$\text{Coverage@}K = \text{Average}\left[\frac{|(\bigcup_{i \in R} Cate_i) \cap (\bigcup_{i \in U} Cate_i)|}{|\bigcup_{i \in U} Cate_i|}\right],$$

与 Diversity@K 相比，这是衡量用户多重兴趣覆盖率的更合理指标，因为后者会偏向于完全随机或不相关的推荐。

3.1.3 基线方法。**FM-BPR** [11] 是一种采用贝叶斯个性化排序准则的因子分解机模型，通过用户-商品交互矩阵学习协同信息。

ContentRec是一种业界广泛使用的基于内容的模型，它从商品标题中学习词袋嵌入，并对用户交互过的物品进行均值池化，以获得用户嵌入。

YouTubeDNN [4] 是业界最受欢迎的深度学习模型之一，它采用神经网络来学习用户/商品嵌入。

BERT4Rec [23] 是一种先进的基于 BERT 的模型，采用双向自注意力架构和掩码训练策略。

3.1.4 实现细节。 我们使用 HuggingFace 发布的 GPT-2 模型 [26]，该模型有 1.17 亿个参数，并使用带权重衰减的 Adam 优化器对模型进行 20 轮微调。我们将学习率设为 0.0001，并使用 2000 个预热步数初始化优化器。我们采用前文所述的方法，在 $k1 \in [0, 3]$, $b \in (0,1)$ 区间内优化 BM25。ContentRec 使用 StarSpace [27] 实现，以学习商品标题嵌入；其他基线方法则通过公开的 GitHub 代码库实现 [2]。对于所有基线方法，我们从 64,128, 256 中选择最佳嵌入维度在验证集上确定，同时按照作者在原论文中的建议设置其余超参数。

3.2 定量分析

3.2.1 整体性能。 表 1 总结了 GPT4Rec 在两个数据集上与不同数量的生成查询及基线方法对比时的 Recall@K 表现。我们提出的框架在两个数据集上均优于所有基线方法：在规模较小的 Beauty 数据集上，Recall@40 相对提升**75.7%**；在规模较大的 Electronics 数据集上，相对提升**22.2%**。

与基线方法的比较表明，商品内容信息和现代语言建模都是取得优异性能的关键因素。一方面，BERT4Rec 借助现代语言建模技术，在基线方法中表现最佳，但将商品视为 ID，未能充分利用商品内容信息。另一方面，ContentRec 使用商品内容信息，并采用词袋嵌入和均值池化建模，仍不足以取得相当的性能。

3.2.2 多查询生成的优势。 观察表 1 中两个数据集的下三角数值，可以发现数值沿行、列和对角线方向呈单调递增趋势。这些趋势表明，多查询束搜索生成策略显著提升了推荐商品的相关性。特别是，生成 K 个查询并为每个查询检索一个物品时，Recall@K的表现最佳。这一发现表明，每个查询都包含足够的细节，能够检索到相关物品。

我们进一步探究推荐商品中用户兴趣的多样性与覆盖率，以及它们与多查询生成策略之间的关系。如表 3 所示，对于 Diversity@K 和 Coverage@K，生成 K 个查询并为每个查询检索一个商品时，效果最佳，这与我们在表 1 中对 Recall@K 的发现一致。类似的单调递增趋势表明，多查询生成策略能够生成更全面的用户兴趣表示。另一方面，正如 Diversity@K 列所示，在不增加查询数量的情况下增加 K，并不会提升多样性。这表明，搜索并不能帮助提升多样性。

[1] https://en.wikipedia.org/wiki/Jaccard_指数 [2] LightFM: https://github.com/lyst/lightfm StarSpace: https://github.com/facebookresearch/StarSpace

DeepMatch：https://github.com/shenweichen/DeepMatch
RecBole：https://github.com/RUCAIBox/RecBole

表3：用户兴趣的多样性和覆盖率与生成的查询数量之间的关系。两个数据集在类别/品牌信息方面的最高值以粗体突出显示。

数据集		美妆（类别）				电子产品（类别）				电子产品（品牌）			
查询数量		5	10	20	40	5	10	20	40	5	10	20	40
Diversity@K	K=5	**0.679**	–	–	–	**0.671**	–	–	–	**0.534**	–	–	–
	K=10	0.654	**0.716**	–	–	0.617	**0.733**	–	–	0.529	**0.643**	–	–
	K=20	0.659	0.706	**0.749**	–	0.605	0.703	**0.778**	–	0.559	0.645	**0.717**	–
	K=40	0.679	0.715	0.749	**0.783**	0.601	0.696	0.762	**0.811**	0.604	0.669	0.724	**0.767**
@K覆盖率	K=5	**0.417**	–	–	–	**0.321**	–	–	–	**0.173**	–	–	–
	K=10	0.472	**0.547**	–	–	0.340	**0.425**	–	–	0.177	**0.239**	–	–
	K=20	0.535	0.602	**0.674**	–	0.364	0.447	**0.537**	–	0.184	0.245	**0.317**	–
	K=40	0.614	0.669	0.726	**0.787**	0.389	0.474	0.562	**0.653**	0.197	0.255	0.324	**0.403**



图 2：一位对美妆产品有多种兴趣的测试用户示例，目标商品以红圈标出。



图 3：一位对电子产品有特定兴趣的测试用户示例，目标商品以红圈标出。

3.3 定性分析

本节通过案例研究，探讨生成的查询在捕捉用户兴趣方面的效果，以及它们在解读用户兴趣方面的作用。我们在图 2 和图 3 中展示了两组测试用户序列及模型输出。

在第一个示例中，测试用户对多个美妆产品类别和品牌有多种兴趣。GPT4Rec 在此情形下能够生成多样化的查询，这些查询不仅涵盖用户历史记录中的部分商品或品牌，还能生成用户历史记录中从未出现过的相关商品，例如“眼影盘”。这表明 GPT4Rec 能够根据用户与商品之间的交互以及商品标题的语义信息，捕捉商品之间的关联。在第二个示例中，测试用户对 Logitech 无线鼠标有非常具体的兴趣。GPT4Rec 成功捕捉到这一特点，所有生成的查询都集中于同一品牌、同一类别，但涉及不同的产品细节。

如上文两个案例研究所示，我们已证明 GPT4Rec 能够生成涵盖用户兴趣不同方面和不同粒度的查询。这些查询直接服务于解读用户兴趣的目的。此外，对这两个示例的比较表明，生成查询的多样性水平会根据用户序列的水平进行调整，这有助于进一步理解用户行为。

4 结论

在这项工作中，我们提出了 GPT4Rec，这是一种新颖的生成式框架，可同时生成个性化推荐并构建可解释的用户兴趣表示。借助先进的语言模型和商品内容信息，GPT4Rec 实现了卓越的性能，并能自然地解决商品冷启动问题等实际难题。所提出的多查询束搜索技术能够从不同方面和粒度生成用户兴趣表示，从而提高推荐结果的相关性和多样性。我们的框架具有灵活性，可整合更先进的生成式语言模型或搜索引擎，以及更出色的生成与检索策略，这些都是值得未来探索的方向。

参考文献

[1] Tom Brown、Benjamin Mann、Nick Ryder、Melanie Subbiah、Jared D Kaplan、Prafulla Dhariwal、Arvind Neelakantan、Pranav Shyam、Girish Sastry、Amanda Askell 等。2020年。语言模型是少样本学习器。《神经信息处理系统进展》第33卷（2020年），1877–1901。[2] Yukuo Cen、Jianwei Zhang、Xu Zou、Chang Zhou、Hongxia Yang 和 Jie Tang。2020年。用于推荐的可控多兴趣框架。载于《第26届ACM SIGKDD知识发现与数据挖掘国际会议论文集》，2942–2951。[3] Minmin Chen、Yuyan Wang、Can Xu、Ya Le、Mohit Sharma、Lee Richardson、Su-Lin Wu 和 Ed Chi。2021年。推荐系统中用户探索的价值。载于《第十五届ACM推荐系统会议论文集》，85–95。[4] Paul Covington、Jay Adams 和 Emre Sargin。2016年。用于 YouTube 推荐的深度神经网络。载于《第10届ACM推荐系统会议论文集》，191–198。[5] Jacob Devlin、Ming-Wei Chang、Kenton Lee 和 Kristina Toutanova。2018年。BERT：用于语言理解的深度双向变换器预训练。arXiv预印本 arXiv:1810.04805（2018年）。[6] Shijie Geng、Shuchang Liu、Zuohui Fu、Yingqiang Ge 和 Yongfeng Zhang。2022年。推荐即语言处理（RLP）：统一预训练、个性化提示与预测范式（P5）。arXiv预印本 arXiv:2203.13366（2022年）。[7] Ruining He 和 Julian McAuley。2016年。起起落落：利用单类协同过滤对时尚趋势的视觉演变建模。载于《第25届万维网国际会议论文集》，507–517。[8] Ashish Jaiswal、Ashwin Ramesh Babu、Mohammad Zaki Zadeh、Debapriya Banerjee 和 Fillia Makedon。2020年。对比自监督学习综述。《Technologies》第9卷第1期（2020年），2。[9] Dietmar Jannach 和 Malte Ludewig。2017年。当循环神经网络遇上邻域：基于会话的推荐。载于《第十一届ACM推荐系统会议论文集》，306–310。[10] Wang-Cheng Kang 和 Julian McAuley。2018年。自注意力序列推荐。载于《2018年IEEE国际数据挖掘会议（ICDM）》。IEEE，197–206。[11] Maciej Kula。2015年。用户与商品冷启动推荐的元数据嵌入。载于《第二届基于内容的推荐系统新趋势研讨会论文集》，该研讨会与第9届ACM推荐系统会议（RecSys 2015）同期举行，奥地利维也纳，2015年9月16日至20日。（CEUR研讨会论文集，第1448卷），Toine Bogers 和 Marijn Koolen（编）。CEUR-WS.org，14–21。http://ceur-ws.org/Vol-1448/paper4.pdf [12] Xuan Nhat Lam、Thuc Vu、Trong Duc Le 和 Anh Duc Duong。2008年。解决推荐系统中的冷启动问题。载于《第2届普适信息管理与通信国际会议论文集》，208–211。[13] Chao Li、Zhiyuan Liu、Mengmeng Wu、Yuchi Xu、Huan Zhao、Pipei Huang、Guoliang Kang、Qiwei Chen、Wei Li 和 Dik Lun Lee。2019年。利用动态路由的多兴趣网络：用于天猫推荐。载于《第28届ACM信息与知识管理国际会议论文集》，2615–2623。[14] Aditya Pal、Chantat Eksombatchai、Yitong Zhou、Bo Zhao、Charles Rosenberg 和 Jure Leskovec。2020年。Pinnersage：用于 Pinterest 推荐的多模态用户嵌入框架。载于《第26届ACM SIGKDD知识发现与数据挖掘国际会议论文集》，2311–2320。[15] Gustavo Penha 和 Claudia Hauff。2020年。BERT 对书籍、电影和音乐了解些什么？探究 BERT 在对话式推荐中的表现。载于《第十四届ACM推荐系统会议论文集》，388–397。[16] Alec Radford、Karthik Narasimhan、Tim Salimans、Ilya Sutskever 等。2018年。通过生成式预训练提升语言理解能力。（2018年）。[17] Alec Radford、Jeffrey Wu、Rewon Child、David Luan、Dario Amodei、Ilya Sutskever 等。2019年。语言模型是无监督的多任务学习器。《OpenAI博客》第1卷第8期（2019年），9。[18] Marc’Aurelio Ranzato、Sumit Chopra、Michael Auli 和 Wojciech Zaremba。2015年。循环神经网络的序列级训练。arXiv预印本 arXiv:1511.06732（2015年）。[19] Steffen Rendle。2010年。因子分解机。载于《2010年IEEE国际数据挖掘会议》。IEEE，995–1000。[20] Stephen Robertson、Hugo Zaragoza 等。2009年。概率相关性框架：BM25及其后续发展。《信息检索基础与趋势》第3卷第4期（2009年），333–389。[21] Andrew I Schein、Alexandrin Popescul、Lyle H Ungar 和 David M Pennock。2002年。冷启动推荐的方法与指标。载于《第25届ACM SIGIR信息检索研究与发展年度国际会议论文集》，253–260。[22] Damien Sileo、Wout Vossen 和 Robbe Raymaekers。2022年。将零样本推荐视为语言建模。载于《信息检索进展：第44届欧洲信息检索研究会议（ECIR 2022）》，挪威斯塔万格，2022年4月10日至14日

’论文集 ，第二部分。施普林格，223–230。

[23] Fei Sun、Jun Liu、Jian Wu、Changhua Pei、Xiao Lin、Wenwu Ou 和 Peng Jiang。2019年。BERT4Rec：基于 Transformer 双向编码器表示的序列推荐。载于第*28*届*ACM*信息与知识管理国际会议论文集。1441–1450。[24] Ashish Vaswani、Noam Shazeer、Niki Parmar、Jakob Uszkoreit、Llion Jones、Aidan N Gomez、Łukasz Kaiser 和 Illia Polosukhin。2017年。注意力机制就是一切。神经信息处理系统进展 30（2017年）。[25] AshwinK Vijayakumar、MichaelCogswell、RamprasathRSelvaraju、QingSun、Stefan Lee、David Crandall 和 Dhruv Batra。2016年。多样化束搜索：从神经序列模型中解码多样化结果。*arXiv*预印本 *arXiv:1610.02424*（2016年）。[26] Thomas Wolf、Lysandre Debut、Victor Sanh、Julien Chaumond、Clement Delangue、Anthony Moi、Pierric Cistac、Tim Rault、Rémi Louf、Morgan Funtowicz 等。2019年。Huggingface 的 Transformers：最先进的自然语言处理。*arXiv*预印本 *arXiv:1910.03771*（2019年）。[27] Ledell Wu、Adam Fisch、Sumit Chopra、Keith Adams、Antoine Bordes 和 Jason Weston。2018年。Starspace：嵌入万物！载于*AAAI*人工智能会议论文集，第32卷。[28] Yongfeng Zhang、Xu Chen 等。2020年。可解释推荐：综述与新视角。基础与趋势® 信息检索 14卷第1期（2020年），1–101。[29] Yuhui Zhang、Hao Ding、Zeren Shui、Yifei Ma、James Zou、Anoop Deoras 和 Hao Wang。2021年。作为推荐系统的语言模型：评估与局限。（2021年）。[30] Kun Zhou、Hui Wang、Wayne Xin Zhao、Yutao Zhu、Sirui Wang、Fuzheng Zhang、Zhongyuan Wang 和 Ji-Rong Wen。2020年。S3-rec：通过互信息最大化进行自监督学习的序列推荐。载于第*29*届*ACM*信息与知识管理国际会议论文集。1893–1902。