

# 【LLM论文阅读】利用 LLMs 进行推荐多样性重排序

原创 方方 方方的算法花园 2024年10月24日 18:09 北京

## 论文标题:

Enhancing Recommendation Diversity by Re-ranking with Large Language Models

论文链接: <https://arxiv.org/pdf/2401.11506>

论文作者所在机构: 爱尔兰科克大学

一句话概括: 首次系统研究通用 LLMs 通过重排提升推荐多样性, 设计多种prompt模板探索其影响, 经多方面实验评估分析, 为推荐系统重排方法选择提供新视角与全面依据。

## 创新点

### 1. 探究 LLMs 在推荐系统多样性重排中的应用

(1) 首次系统地研究了通用大型语言模型 (LLMs) 在推荐系统中通过重排提高推荐多样性的能力, 拓展了 LLMs 在推荐系统领域的应用范围, 为解决推荐多样性问题提供了新的视角和方法。

(2) 通过实验证明了 LLMs 能够理解重排任务并在一定程度上提高推荐多样性, 尽管与传统重排方法相比在相关性指标上存在差距, 但具有一定潜力。

### 2. 设计多样化的prompt模板

(1) 设计了八种不同的prompt模板, 涵盖了平衡相关性与多样性以及仅最大化多样性等不同目标, 这些模板具有不同的复杂度, 包括简单的高级任务描述和更复杂的带有额外信息及重排指南的模板, 以探索不同方式对生成结果的影响。

(2) 研究发现不同提示模板对重排性能有显著影响, 且特征感知模板在平均性能上表现更好, 这为后续基于 LLMs 的重排研究中提示模板的设计提供了有价值的参考。

### 3. 全面的实验评估与分析

(1) 在两个公开数据集 (动漫电影和书籍领域) 上进行了广泛的实验, 对比了基于 LLMs 的重排方法 (如 ChatGPT、InstructGPT、Llama2 - 7B - Chat 和 Llama2 - 13B - Chat) 与传统贪婪多样化重排方法及随机重排方法的性能。

(2) 除了传统的准确性 / 多样性性能指标外, 还深入分析了不同方法在成本、对数据的控制、内存占用、处理冷启动问题能力、生成有效排名能力以及在不同推荐领域适用性等多方面的权衡, 为推荐系统提供者在选择重排方法时提供了全面的决策依据。

## 具体方案

### 1. 初步研究

(1) **重排任务测试:** 设计简单prompt, 让 LLMs (如 ChatGPT 和 Llama - 2 - Chat) 对 20 个动漫电影标题列表进行重排, 任务包括按字母顺序、流行度和发布日期排序, 以探究 LLMs 是否能执行重排任务及在不同任务难度下的表现。

**(2) 多样性理解测试：**设计提示让 LLMs 比较两个各含 10 个动漫标题的列表的多样性，且不提供多样性定义，以判断 LLMs 对物品多样性的理解能力。

2. prompt模板设计

**(1) 模板框架：**将重排任务框架为文本生成任务，设计通用prompt模板，包含提供上下文和指令的第一部分、规定输出格式的第二部分以及候选列表的第三部分。

```
<Instructions>:
You are given a ranked recommendation list of {m} items for a user,
delimited by triple backticks.
Your task is to re-rank this candidate list and provide a final top-{n}
recommendation list where the goal is to {goal string}. Strictly use the
following format for the output, and don't provide additional information.

<Output format>:
1-> <item name>
2-> <item name>
...
{n}-> <item name>

<CL>:
...
1. {candidate item name ranked #1 in candidate set} {additional_features}
2. {candidate item name ranked #2 in candidate set} {additional_features}
...
{m}. {candidate item name ranked #m in candidate set} {additional_features}
...
```

Prompt 1. Generic re-ranking prompt template. Tags in angular brackets delimit sections (but they do not appear in the final prompt), and placeholders for data are enclosed in braces.

**(2) 模板类型与目标：**设计八种模板类型（T1 - T8），通过不同的重排目标（如平衡相关性与多样性或仅最大化多样性）和指令映射不同的lambda值；部分模板为特征感知模板（T5 - T8），通过提供物品的描述等额外信息增强 LLMs 重排能力。

Table 1. Mapping of placeholders in the generic prompt template.

Template id	$\lambda$	{goal_string}	{additional_features}
T1	0.5	"balance relevance and diversity"	None
T2	0	"maximize the items' diversity in the list"	None
T3	0	"maximize the items' genre-based diversity in the list"	None
T4	0.5	"balance relevance and genre-based diversity"	None
T5	0.5	"balance relevance and diversity"	List of the candidate item's genres
T6	0.5	"balance relevance and genre-based diversity"	List of the candidate item's genres
T7	0.5	"balance relevance and diversity. Guidelines to perform the re-ranking are: Use the plot summary information of each item attached in curly bracke"	Candidate item's plot summary
T8	0	"maximize the books' diversity in the list. Guidelines to perform the re-ranking are: Use the plot summary information of each item attached in curly bracket"	Candidate item's plot summary

3. 实验设置与实施

**(1) 数据集与预处理：**使用 Anime Recommendation Database 和 Goodreads Book Graph Datasets 数据集，对其进行预处理，包括移除未知体裁的物品、处理动漫电影标题格式、过滤用户评分数量、映射评分尺度等操作。

**(2) 实验流程：**将数据集随机分割为训练集和测试集，训练 Matrix Factorization 基线推荐器（MF），为测试用户计算候选物品列表，使用不同的 LLMs（OpenAI 的 ChatGPT、InstructGPT 和 Meta 的 Llama2 - Chat 系列模型）基于设计的提示模板对候选物品列表进行重排，同时采用传统的 MMR、xQuAD、RxQuAD 贪婪重排方法和随机重排方法作为对比，在处理 LLMs 输出时，对部分无效输出进行智能修正或填充。

(3) **超参数调整**: 使用验证集调整基线推荐器 MF 的因子数量 $k$ 、最终推荐列表大小 $n$ 和候选列表大小 $m$ 等超参数。

(4) **评估指标**: 采用 Precision、Recall 和 NDCG 评估推荐的相关性, 使用 $\alpha$ -NDCG、Intra - List Diversity (ILD)、Expected Intra - List Diversity (EILD)、Subtopic - Recall (SRecall) 和 relevancy - aware SRecall (rSRecall) 评估推荐多样性。

#### 4. 结果分析与讨论

(1) **LLMs 与其他重排方法对比**: 分析 LLMs 基重排器与随机重排器和传统重排器在相关性和多样性指标上的差异, 发现 LLMs 能提高多样性但牺牲了一定相关性, 且在相关性感知指标上劣于传统重排器, 同时比较不同 LLMs 模型 (如 ChatGPT 表现较好, Llama2 - 13B - Chat 表现较差) 和不同数据集上的性能差异, 并分析原因 (如抽取低排名物品、随机推荐比例等因素影响)。

(2) **不同提示模板对比**: 研究不同提示模板对 LLMs 基重排器性能的影响, 发现特征感知模板总体有益, 但不同模板在不同数据集和指标上表现各异, 模板设计需根据具体情况定制。

(3) **重排成本分析**: 对比不同重排方法的成本, 包括推理时间和财务成本 (API 调用费用等), 发现传统重排方法在速度和资源需求方面优于 LLMs 基重排方法, 同时指出 LLMs 成本下降趋势可能改变这种局面。

#### prompts示例

(1) 评估 ChatGPT 是否可以按字母顺序重新排列动画标题列表

##### A.1 Preliminary study

Q1-P1. This prompt assesses whether ChatGPT can re-rank a list of anime titles into alphabetical order.

You are given a list of 20 anime, delimited by triple backticks.  
Your task is to re-rank this list in alphabetical order.  
...

0.-> Hundred  
1.-> The Millionaire Detective - Balance:UNLIMITED  
2.-> Freaky Highschool  
3.-> Arte  
4.-> Infinite Ryvius  
5.-> Is this A Zombie? of the Dead  
6.-> A Christmas Adventure  
7.-> Junjo Romantica  
8.-> Rokka:Braves of the Six Flowers  
9.-> Wooser's Hand-to-Mouth Life:Awakening Arc  
10.-> Jormungand Season 2:Perfect Order  
11.-> Genshiken  
12.-> Shirobako  
13.-> Scum's Wish  
14.-> Inuyashiki:Last Hero  
15.-> My Name Is Teppei  
16.-> Magical Project S  
17.-> Hand Shakers  
18.-> Tekkaman Blade  
19.-> Pretty Cure  
...

(2) 评估 ChatGPT 是否可以按受欢迎程度对动画标题列表进行重新排名

Q1-P2. This prompt assesses whether ChatGPT can re-rank a list of anime titles by popularity.

You are given a list of 20 anime, delimited by triple backticks.  
Your task is to re-rank the anime in this list by popularity.  
...  
Tsubasa RESERVoir CHRoNiCLE Season Two  
El-Hazard:The Wanderers  
Gintama Season 2  
Beet the Vandel Buster  
Umamusume:Pretty Derby  
Bobobo-bo Bo-bobo  
GG Bond:Ultimate Battle Prequel  
Rozen Maiden:Zuruckspulen  
Cardfight!! Vanguard:Shinemon Arc  
Prince Mackerroo  
Samurai Deeper Kyo  
New Game!!  
Trigun  
ChaoS;Child  
Beautiful Bones -Sakurako's Investigation-  
The Ones Within  
Tiger & Bunny  
Cat Eyed Boy  
Vampire Princess Miyu  
Show By Rock!! 2  
...

(3) 评估 ChatGPT 是否可以按发布日期对动画标题列表进行重新排名。

Q1-P3. This prompt assesses whether ChatGPT can re-rank a list of anime titles by release date.

You are given a list of 20 anime, delimited by triple backticks.  
Your task is to re-rank anime in this list by release date (from newest to oldest).  
...  
Digimon Adventure 02  
Chobits  
Fighting Spirit  
Chrono Crusade  
The Great Passage  
Kiba  
Green Green  
PocoColies  
Sekko Boys  
Teasing Master Takagi-san  
Steel Angel Kurumi  
A Centaur's Life  
Persona:Trinity Soul  
Magical Princess Minky Momo:Hold on to Your Dreams  
Coral Investigates  
The Pet Girl of Sakurasou  
The Affectuous Family  
Earl and Fairy  
Vividred Operation  
Golden Time  
...

(4) 评估 ChatGPT 是否可以选两个动画标题列表中更多样化的一个

Q2-P1. This prompt assesses whether ChatGPT can select the more diverse of two lists of anime titles.

---

You are given two lists of 10 anime each, list A and list B, delimited by triple backticks.

Which of the two lists is more diverse?

```

LIST A:

Mr. Stain on Junk Alley

Let's Dance with Papa

King of Classroom, Yamazaki

Poo-Neko

The Adventures of the Little Koala

Oscar's Oasis

The Affectuous Family

Little Wansa

Hippo and Thomas

Weekly Shimako

LIST B:

Hundred

The Millionaire Detective - Balance:UNLIMITED

Freaky Highschool

Arte

Infinite Ryvius

Is this A Zombie? of the Dead

A Christmas Adventure

Junjo Romantica

Rokka:Braves of the Six Flowers

Wooser's Hand-to-Mouth Life:Awakening Arc

Jormungand Season 2:Perfect Order

```

---

#### (5) 来自模板T1并使用动画数据集的prompt示例

You are given a ranked recommendation list of 40 items for a user, delimited by triple backticks.

Your task is to re-rank this candidate list and provide a final top-10 recommendation list where the goal is to balance relevance and diversity.

Strictly use the following format for the output, and don't provide additional information.

1-> <anime name>

2-> <anime name>

...

10-> <anime name>

```

1. Dragon Ball Z

2. Naruto

3. Deadman Wonderland  
...  
9. Code Geass:Lelouch of the Rebellion R2  
10. One Punch Man  
...  
15. Haikyu!!  
...  
22. Made in Abyss  
...  
27. Death Parade  
...  
30. Attack on Titan Season 3 Part 2  
...  
32. Re:ZERO -Starting Life in Another World  
...  
34. The Promised Neverland  
35. Demon Slayer:Kimetsu no Yaiba  
...  
40. Your Lie in April  
...

END

LLM与推荐 15    LLM论文阅读 13    #LLM学习 12

LLM与推荐 · 目录

上一篇

【LLM论文阅读】LlamaRec:具有高效检索与排序的两阶段推荐框架

下一篇

【LLM论文阅读】DMPO 与 S-DPO: LLM推荐排序模型直接偏好优化研究