

赞同 27

分享

## META 2023 - 基于图模型的用户个性化兴趣探索



SmartMindAI

专注搜索、广告、推荐、大模型和人工智能最新技术，欢迎关注我

已关注

27 人赞同了该文章

### 论文原文《PIE: Personalized Interest Exploration for Large-Scale Recommender Systems》

#### Introduction

探索策略对于大规模**推荐系统**<sup>+</sup>非常重要，本文提出了一个框架用于在Facebook Watch的真实场景中探索，这个框架可以轻松地与现有的部署系统集成。

本文分为三部分：

- 个性化探索空间：本文使用个性化PageRank (PPR) 算法，一种基于随机游走的**非参数方法**<sup>+</sup>，在图中的每个节点中找到一组相邻节点，以发现与用户有高亲和力的新创作者，实现个性化探索空间。
- 在线探索框架，针对遍历不确定兴趣空间导致用户相关性降低的问题，提出将其表述为在线强盗问题，后面探讨实现方法，以提高在线探索效率并减少用户参与损失。
- 平衡探索/利用的交付：探讨了如何在用户信息流中结合探索性内容，提出了特殊的序模块。我们讨论了一种方法，能结合相关性和探索价值，以优化长期奖励。

Facebook Watch是全球最大的视频内容发现和分享平台，月观众超12.5亿。其个性化**推荐引擎**<sup>+</sup>依据用户兴趣、新趋势、创作者亲和力等因素推荐视频内容，设计类似两阶段推荐系统，包括检索相关候选项目和对项目进行排序。

#### Related Work

近年来，通过从观察到的交互数据中提取用户偏好来为用户提供个性化内容的推荐系统非常成功。然而，这样的系统往往会强化或支持用户当前的偏好，并往往会加强流行度偏见。最近的研究开始关注反馈循环和长期影响，包括曝光不平等、极化和**回声室效应**<sup>+</sup>。探索推荐系统对多样性和新颖性的需求以及解决过滤器泡沫、回声室和极化问题至关重要。

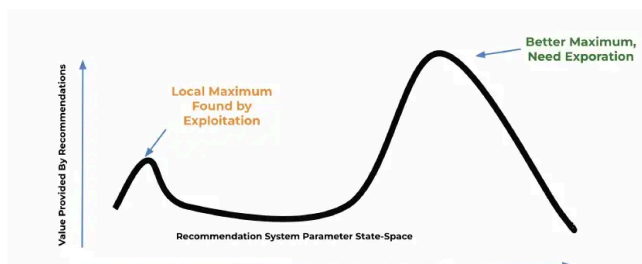


图 1: Exploration of users' interests could lead to a better maxima on the value provided by recommender systems.

推荐系统模型通常依赖数据来准确解决问题。然而，如果没有数据，这些系统在寻找新的兴趣或逃离局部最小值<sup>+</sup>方面的能力可能会受到限制。本文讨论了推荐系统模型在数据缺乏的情况下，寻找新兴趣和避免局部最小值方面的能力可能受到限制。为了解决这一问题，推荐系统需要系统地探索各种内容，以便为用户提供更相关的建议。

本文探讨了仅利用对用户理解会导致内容重复，从而错过用户的其他兴趣的问题。为了找到更多或为用户引入新的兴趣，我们希望探索其他不太确定的主题。这种探索需要高效、有效，并与开发策略保持最佳平衡。

## System Design

我们在现有的推荐系统上设计了探索系统，与现有推荐系统交互，仅在排名后的混合阶段，将最佳探索视频与非探索视频混合，易于应用到任何现有推荐系统上。如图蓝色路径所示。

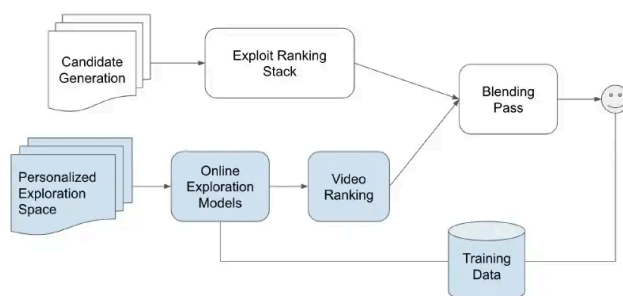


图 2: Incorporating personalized exploration in existing recommender system ranking framework. 知乎 @SmartMindAI

## Personalized Exploration Space

探索每个用户与每个内容之间的关联性可能非常耗时。相反，我们尝试学习每个用户与每个属性之间的关联性，以替代探索每个用户与每个内容之间的关联性，降低集合大小，从而节省时间。

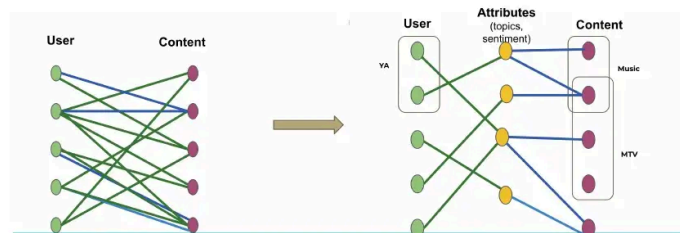


图 3: Moving from user-content to user-attribute-content 知乎 @SmartMindAI

我们使用个性化PageRank (PPR) 算法创建了用户-创作者亲和力。该算法基于随机游走，利用多跳信息，对低参与度创作者友好。方法：在用户和创作者之间创建一个双向图，互为源和目标。然后，我们根据用户过去几天的参与度为边分配权重。个性化创作者排名，过滤出创作者节点，获得相似创作者集。聚合前k个创作者。添加新颖性和质量过滤器，提高探索效率。为了提高探索效率，我们在创作者级别添加了两个过滤器。新颖性过滤器会移除用户之前已经交互过的创作者，确保探索系统只捕捉用户的新兴趣。质量过滤器会移除低质量的创作者，例如具有“诱导观看”内容的创作者或通过探索获得的参与度较低的创作者。

## Online Exploration

用户探索的目标是通过向用户推荐新颖的创作者来减少不确定性。这通过从大量选项开始，然后收敛到一个状态，其中一个新颖的创作者被认为“过期”或“已连接”。虽然此过程存在机会成本，但用户可能会从更多样化的推荐中受益。为用户创建的空间探索过程可以表述为在线强盗问题。一个简单的解决方案可能是汤普森采样，而更完善的解决方案可能是考虑用户、创作者和交互特征

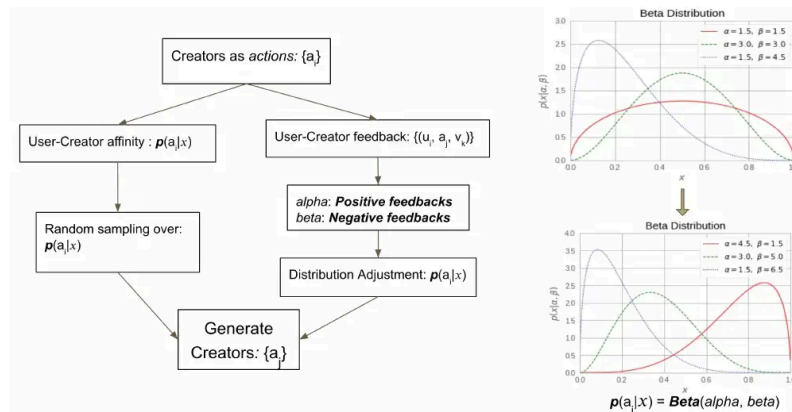


图 4: (Left) Random sampling vs Thompson sampling (Right) Distribution of sampling creators evolving over time

### Delivery to balance explore/exploit

探索组合: 我们使用了一个概率混合层, 它考虑了用户当前会话中的探索组合, 并插入了一个探索视频来强制执行预定的组合。

## Experiment

### Setup

为了测试我们提出的方法, 我们在Facebook Watch 上进行了一项在线实, 分为四组用户。条件是是否提供探索内容及推荐系统是否使用探索数据进行训练。表显示组合情况。

### Metrics

在理解视频探索的价值时, 最大的挑战之一是找到合适的指标。常用的指标, 如观看时间, 可能会随着探索利用机会成本<sup>+</sup>来展示更高不确定性的内容而暂时恶化。我们定义了一组指标作为这些联系的代理。

创作者联系 (SCC) :使用创作者作为偏好/内容的代理, 因为创作者通常会发布具有一致风格、主题等内容。

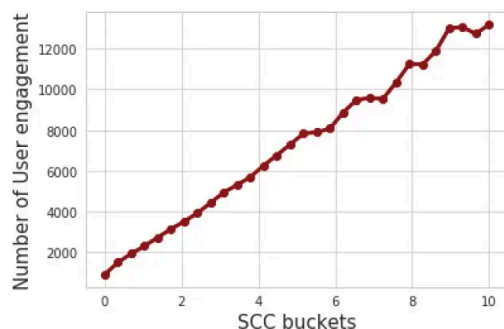


图 5: SCC showing strong correlation with future engagement

## Results

加了0.26%，新颖SCC增加了0.85%，而对用户的视频消费或留存没有统计上的显著影响。

**\*\*系统探索价值\*\***：第3组和第1组比较，反映探索数据对推荐系统的益处。参与度提高0.28%，SCC无显著变化。 **\*\*严格探索价值\*\***：第2组与第1组比较，显示增加探索内容的影响。参与度下降0.53%，但用户创建者连接的SCC增加0.55%。

本文研究了探索对用户参与度的影响，通过计算每个子话题的用户参与度并绘制[直方图](#)，发现探索能够平衡印象和参与度的分布。

Model Description	SCC	SCC DAU	NOVEL SCC	Engagement
User Exploration Value	+3.50%	+0.26%	+0.85%	Neutral
System Exploration Value	Neutral	Neutral	Neutral	+0.28%
Strict Exploration Value	+0.55%	Neutral	+0.11%	-0.53%

表 2: Exploration metrics impact in A/B test relative to production. 知乎 @SmartMindAI

Conclusion

在本文中，我们提出了一个用于大规模推荐系统探索的框架，该推荐系统服务于数十亿用户。该框架可以很容易地集成到现有的推荐系统中

为了捕捉探索的价值，本文定义了一组代理指标，用于衡量用户与创建者之间的有意义联系，并进行了广泛的A/B测试，以了解整体参与度和SCC指标收益。通过分析，展示了探索方法如何通过与新创建者建立联系来帮助捕捉用户的新兴趣。算法已部署在Facebook Watch的生产环境中。

跟踪最新技术，点击关注我  
[www.zhihu.com/people/smartmindai](http://www.zhihu.com/people/smartmindai)



编辑于 2024-02-18 22:20 · IP 属地北京

[Meta分析](#) [推荐系统](#) [Meta \(Facebook\)](#)



理性发言，友善互动



还没有评论，发表第一个评论吧