

## • ABSTRACT

- 基于历史行为学习用户描述是现代推荐系统的核心。
- 序列推荐的最新进展令人信服地展示了从给定行为序列中提取有效用户描述的高能力。
- 当前问题
  - 由于用户的历史记录存在交互稀疏性以及其中存在的噪声，仅对观察行为序列进行建模可能最终导致系统脆弱且不稳定。
- 解决方案
  - 作者建议通过对反事实数据分布进行建模来学习准确和稳健的用户表示，这些表示需要对（攻击）噪音行为不那么敏感，让其更多地信任不可或缺的行为。
  - 具体来说，给定观察到的行为序列，提议的 CauseRec 框架在细粒度项目级别和抽象兴趣级别上能够识别可有可无的项目和不可或缺的概念。
    - CauseRec通过替换原始概念序列中的可有可无和不可或缺的概念，从反事实数据分布中有条件地对用户概念序列进行采样。
    - 利用从合成的用户序列中获得的用户表示，CauseRec通过对比反事实和观察结果来进行对比用户表示学习。

## • INTRODUCTION

- 当前问题
  - 噪声
    - 由于无处不在的干扰因素可能会影响用户的第一印象，用户的兴趣与其点击行为之间存在不一致，称为自然噪声。
  - 数据稀疏
    - 在推荐系统中，与在大型实时系统中可以轻松达到 1 亿的项目库相比，用户通常只与有限数量的项目进行交互。
    - 仅对既稀疏又有噪声的观测行为序列进行建模，最终可能会得到一个不太令人满意的脆弱系统。
- 解决方案
  - 作者提出了一种用于顺序推荐的反事实用户序列合成，简称为CauseRec。
  - CauseRec 面对数据稀疏问题的本质是对反事实数据分布进行建模，而不是对观测稀疏数据分布进行建模，后者可以是前者的子集。
    - 具体来说，给定观察到的行为序列，作者在细粒度项目级别和抽象兴趣级别识别不可或缺/可有可无的概念（概念表示用户兴趣/偏好的某个方面）。
    - 通过修改可有可无的概念得到反事实的正用户表示，通过替换必要的概念得到反事实的负用户表示。
    - 使学习的表示变得稳健，因为它们对可有可无的/嘈杂的概念不那么敏感。

- 学习到的表述可能更准确，因为他们更信任更能代表用户兴趣的不可或缺的概念。

- 总结

- 本文主要针对的是如何从用户历史行为序列中学习更鲁棒更有效的表征。
- 用户历史行为序列中是存在噪声的，并且序列本身是稀疏的，因此本文针对现存的问题，提出CauseRec方法通过对反事实数据分布进行建模，更多地信任不可或缺的行为从而提高鲁棒性。
- 给定行为序列，CauseRec识别可有可无的概念和不可或缺的概念，CauseRec通过替换原始序列中的可有可无和不可或缺的概念，有条件地从反事实数据分布中采样用户概念序列。