CE-NR

- ABSTRACT
  - 当前困难
    - 虽然神经推荐器近年来已成为最先进的技术，但深层模型的复杂性仍然使得为最终用户生成有形解释成为一个具有挑战性的问题。
    - 现有的方法通常基于对各种特征的注意力分布，这些特征作为解释的适用性仍然存在问题，并且对于最终用户来说难以掌握。
  - 解决方案
    - 基于一小部分用户自己行为的反事实解释已被证明是有形性问题的可接受解决方案。
    - 然而，目前关于这种反事实的工作不能轻易地应用于神经模型。
    - 提出了ACCENT，这是为神经推荐者寻找反事实解释的第一个通用框架。
- INTRODUCTION
  - 解释最先进神经推荐的典型方法面临某些问题：
    - 他们经常依靠注意力机制来寻找重要的词、评论或图像中的区域，这仍然存在争议。
    - 使用用户和项目 之间的连接路径，这些路径可能不是真正可操作的并且有隐私问题。
    - 他们使用外部项目元数据，例如评论或图像，这些元数据可能并不总是可用。
  - 反事实解释是一组用户自己的行为，当移除这些行为时，会产生不同的推荐（替换项）。
  - 伴随着大量的解释工作，推荐模型本身也变得越来越复杂。
- 总结
  - 问题
    - 深层神经推荐模型的复杂度影响了模型的可解释性。
    - 现有的解决方法为基于对各种特征的注意力分布，但是这些特征作为解释的适用性仍然存在问题，并且对于最终用户来说难以掌握。
  - 解决方案
    - 作者提出使用反事实解释来提升深层神经推荐模型的可解释性，并提出了一个为神经推荐模型寻找反事实解释的第一个通用框架ACCENT。