

• ABSTRACT

- 各种可解释模型的评估方法因模型而异，这使得比较不同模型的可解释性非常困难。
 - 为了实现评估推荐解释的标准方法，作者提供了三个用于解释排名的基准数据集（注释为EXTRA），在这些数据集上，可通过面向排名的指标来衡量解释性。
 - 构建这样的数据集带来了巨大的挑战。
 - 在现有的推荐系统中，用户项解释三元组交互作用很少见，因此如何找到替代方案成为一个挑战。作者的解决方案是从用户评论中找出几乎相同的句子。
 - 如何有效地将数据集中的句子分类为不同的组，它使用二次运行时复杂性来估计任意两个句子之间的相似性。
- 为了缓解这个问题，作者提供了一种基于局部敏感哈希（LSH）的更有效的方法，该方法可以在给定查询的次线性时间内检测到近似重复。

• INTRODUCTION

- 可解释的推荐系统 (XRS) 不仅为用户提供个性化推荐，而且还证明了他们被推荐的原因，已成为近年来的一个重要研究课题。
 - 与旨在解决用户信息过载问题的协作过滤 (CF) 和协作推理 (CR) 等其他推荐算法相比，XRS 进一步提高了用户的满意度和通过帮助他们更好地理解推荐项目来体验。
 - 实际上，解释和推荐本身一样重要，因为通常没有绝对的“对”或“错”来推荐哪些项目。
 - 相反，用户可能对多个项目都感兴趣，这完全取决于我们如何向用户解释我们的建议。
- 然而，由于解释可以采用各种形式，例如预定义的模板、生成的文本或知识图谱上的决策路径，有时很难评估不同方法产生的解释。
- 在这项工作中，为了使可解释推荐的标准评估成为可能，作者提出了三个基准数据集，可以通过标准排名指标（例如 NDCG、Precision 和 Recall）对推荐解释进行定量评估。
- 解释排序的想法受到信息检索的启发，信息检索可以智能地为给定查询的可用内容（例如，文档或图像）排序。
 - 在作者之前关于解释生成的工作中，其发现大量生成的句子是训练数据中常见的句子，例如，“the food is good”作为对推荐餐厅的解释。这意味着生成模型正在拟合给定的样本，而不是创建新的句子。
 - 此外，即使是强大的语言模型，例如在大型文本语料库上训练的 Transformer，也可能生成偏离事实的内容，例如“四角独角兽”。
- 因此，作者创建了三个解释排名数据集（表示为额外数据集），用于可解释的推荐研究。
 - 它们建立在用户生成的评论之上，这些评论是用户对项目的真实反馈的集合，因此是解释的理想代理。
 - 此外，当新发布的评论包含新的项目特征或最新的表达时，新的解释可以进一步丰富数据集。

- 然而，简单地采用评论或其句子作为解释不太合适，因为两个评论/句子不太可能完全相同，因此，每个评论/句子在数据中只出现一次。
 - 出于这个原因，几乎没有用户共享相同的评论/句子，这使得设计用于解释排名的协作学习算法变得困难。
 - 此外，并非用户评论中的所有句子都具有解释目的。
- 解决方案
 - 从不同评论中共同出现的评论中提取解释句子，以便将不同的用户-项目对与一个特定的解释联系起来，并构建用户-项目-解释三元组交互。
 - 通过查找常用语句，可以保证解释的质量，如可读性和表达能力。
 - 这种类型的文本解释可以非常有效地帮助用户做出更好、更明智的决定。
- 后续问题
 - 如何在数据集中的评论中检测几乎相同的句子。
 - 在这种情况下，数据聚类是不可能的，因为它的质心数是预定义的和固定的。
 - 计算数据集中任意两个句子之间的相似度很实用，但效率较低，因为它具有二次时间复杂度。
- 解决方案
 - 作者开发了一种方法，可以将句子分为不同的组，基于为近似重复检测而设计的局部敏感哈希（LSH）。
 - 此外，因为有些句子不太适合解释目的，所以作者只保留那些同时包含名词和形容词的句子，而不保留人称代词，通过这种方式可以获得高质量的解释，这些解释是以特定的观点谈论项目特征，而不是通过个人经验。
 - 在整个过程之后，解释语句仍然是个性化的，因为它们类似于传统推荐的情况，其中具有相似偏好的用户会写出几乎相同的评论语句，而相似的项目可以用相同的解释来解释。
- 总结
 - 作者构建了三个由用户-项目-解释交互组成的大型数据集，可以通过标准排名指标（例如 NDCG）评估其可解释性。数据集和代码可在线获取。
 - 作者在创建此类数据集时解决了两个关键问题。
 - 解释和用户/项目之间的交互。
 - 对几乎相同的句子进行分组的效率。