

• ABSTRACT

- 在人工推荐设置中，人们经常使用软属性来表达对项目的偏好，包括诸如电影情节的原创性、场地的嘈杂或食谱的复杂性等概念。
- 虽然在推荐系统的上下文中广泛研究了二进制标记，但软属性通常涉及主观和上下文方面，不能以这种方式可靠地捕获，也不能在知识库中表示为客观的二进制事实。
- 作者提出了一种更自然的表示，作为个性化的相对陈述，而不是绝对的项目属性。
- 并提出了新的数据收集技术和评估方法，以及一个新的公共数据集。
- 还提出了一套评分方法，从无监督到弱监督再到完全监督，作为解释和处理基于软属性的评论的一个步骤。

• INTRODUCTION

- 推荐系统根据用户的偏好向用户提供个性化的产品或服务推荐。
 - 推荐系统的一个关键要素是用户能够提供关于他们喜欢和不喜欢的反馈。
 - 这本质上是一个顺序交互过程，并且越来越倾向于设计自动化系统，该系统通过多轮对话来征求用户的偏好并提供个性化建议。
 - 理想情况下，这样的系统应该能够结合任何自然语言用户对项目的反馈。
- 虽然软属性在自然对话中很常见，但解释它们（即确定它们应用于项目的程度）代表了作者要解决的问题。
 - 相比之下，过去大多数基于属性的评论工作都是用专注于二元和明确属性的数据集进行训练和评估的。
 - 关键困难源于软属性不受约束的自然语言，即没有普遍认同的属性，或者不同的人有不同的规范、期望和阈值，这与经过充分研究的客观属性形成鲜明对比。
 - 由于其主观性，在知识库中将软属性表示为属性-值对容易出错，而且通常不可行。
 - 关键的是，虽然软属性类似于社会标签，但至少有两个关键差异使得应用标签集合具有挑战性。
 - 标签是二进制标签，不允许在相关项目中进行相对排序。因此，大多数基于标签的评估侧重于比较有和没有标签的项目，对细粒度结构不敏感。
 - 虽然原则上标签和软属性都是不受约束的自然语言，但大多数标签方法可能会倾向于使用一致的词汇，例如显示频繁的去标签或预先定义流派本体。
- 具体而言，本文讨论了软属性的解释问题。
 - 与将主观项目描述合并到潜在用户偏好表示中以提高端到端推荐性能（以成功率或对话轮数衡量）的任务不同。
 - 相反，作者的目标是能够明确衡量软属性适用于给定项目的程度，以及量化软属性的“柔软度”（即主观性）程度。
- 三个主要贡献

- 作者开发了一个可重用的测试集合，包括一组软属性和与该属性（针对特定用户）相关的真实项目排序，以及一个评估指标。用以研究的软属性来自电影偏好的真实对话。
 - 通过参与者临时提到的抽样属性，作者开发了一种新颖的受控多阶段人群标记机制来收集个性化偏序的真实数据，同时保持了人员的低认知负荷。
 - 将属性与标签进行对比，作者注意到其研究的软属性中只有 47% 作为标签出现在常用的 MovieLens TagGenome 中。
 - 为了评估评分函数的质量，作者还基于与结构化真实数据排名的一致性，对已建立的排名相关性提出了一种新的加权扩展。
- 对软属性的主观性或“软度”的明确量化。
 - 作者确定了区分更多与较少主观软属性的方法，并衡量这如何影响项目评分。
 - 这一结果也对标准标签产生了影响，引起了对主观性作用的关注。
- 作者解决了基于软属性的评估问题，并提供了经验证据来证明无偏差收集真实数据的重要性。
 - 特别是，基于软属性的评估是相对于给定软属性对给定锚项对项目进行排名的任务。并为此任务设计了三类方法（无监督、弱监督和完全监督），它们对应于越来越先进的方法。
 - 这些方法在两个测试集合上进行了比较：一个是基于现有的社交标签创建的，另一个是使用作者提出的方法构建的。
 - 两个测试集的结果之间存在差异，表明基于标签的测试集对项目排名改进视而不见，这使得该领域的进展变得困难。
 - 这证明了新的数据集和评估方法如何能够在对自然语言批评的细微解释方面取得重大进展，而这在以前是无法可靠解决的。
 - 作者还分析了关于属性“软度”的性能，发现方法在具有较高一致性的属性上的表现明显优于那些具有较低一致性的属性。

• 总结

- 主要新颖之处在于软属性概念的形式化，这为与对话推荐系统更自然的交互开辟了新的可能性。
- 主要技术贡献有三方面：
 - 提出了一种有效的方法，用于根据给定的软属性比较项目并发布新的数据集。
 - 引入了量化软属性主观性的度量，并进行了“软度”分析。
 - 引入并形式化了一个基于软属性的更精细的评估任务，开发了无监督、弱监督和完全监督的方法。