TTES

- ABSTRACT
  - 当前困难
    - 缺乏基本事实解释使得评估可解释人工智能系统（XAI）（如可解释搜索）变得困难。
  - 解决方案
    - 作者提出了一个可解释的搜索系统，重点是评估XAI在数据搜索方面的可信度以及性能。
  - 介绍
    - SIMFIC 2.0（小说中的相似性）
      - 该系统检索与设置中所选书籍范例相似的书籍。其动机是解释小说中的相似性概念。
      - 作者为小说手工提取的可解释特征，并通过拟合线性回归和基于相似性度量的局部解释来提供全局解释。
      - 可信度方面使用用户的调查报告进行评估，而排名性能则通过用户的点击数据进行比较。
      - 眼球追踪用于用户在与界面交互时对解释元素的注意力调查。
      - 最初的实验显示了在系统可信度方面显著的统计结果，为正在被调查的有趣研究方向铺平了道路。
- INTRODUCTION
  - 可解释人工智能（XAI）系统试图“解开”人工智能系统决策背后隐藏的逻辑。
    - 目的
      - 向最终用户提供透明度，并获得用户信任。这通常是一项监管要求，如欧盟GDPR“解释权”。
    - XAI 根据不同的使用场景（分类或检索）有着不同的判断。
      - 在分类设置中，重点通常是开发附加方法，如 LIME、LRP来解释分类决策。
      - 在 IR 设置中，重点是解释相对排名和全局排名。
      - 在搜索书籍的上下文中，目标是向最终用户解释文本中的相似性概念。
      - 从人文角度来看，书籍之间的相似性可能取决于主观方面，例如写作风格、情感和相关方面。
    - 问题
      - 由于这种判断经常发生在 IR 系统中，因此在获得真实数据相关性方面存在挑战。
      - 当没有基本事实解释时，该问题在 XAI 的设置中会更加严重。
    - 解决方案
      - 作者试图将重点放在评估方面以改进现有的XAI搜索。
  - SIMFIC 1.0引入了一个可解释的图书搜索系统，作为寻找小说书籍的一种新方法。
    - 每本书都由一个紧凑的（22 个特征）、手工提取的可解释特征空间表示。

    - 通过比较图书特征向量，计算用户选择的查询图书与所有其他图书之间的相似性。
    - 它基于使用分类器的特征选择，为主页上排名前十的书籍提供了一个全局级别的解释。
    - 检索场景被构建为二元分类设置，其中前十本相关书籍属于第一类，而语料库中的所有其他书籍都属于第二类。
        - 全局解释显示了使排名前十的书籍与其他书籍不同的关键特征。
    - 当将排名与基于词袋特征表示的检索模型进行比较时，SIMFIC 1.0 给出了令人期待的结果。
    - 缺点
        - 没有为每本检索到的书提供局部解释，并且没有评估特定XAI方面的解释。
- 总结
    - SIMFIC 2.0 的目标是通过专注于 XAI 评估来解决这些不足。
    - 生成局部解释
        - 作者开发了生成局部解释的方法，探索相似性度量。
        - 作者通过将问题作为线性回归设置而不是二元分类来探索一种新的全局解释方法。
    - 对其他语言的概括
        - 研究对德语书籍整体方法的概括。
    - XAI 评估
        - 作者通过调整 XAI 社区的现有定义来评估 XAI 评估的特定方面，例如用户研究中的可信度。
    - 排名评估
        - 作者通过分析用户点击数据来比较检索性能。