

Survey on Complex Knowledge Base Question Answering: Methods, Challenges and Solutions

综述动机

以前的综述文章仅从技术的角度提供了方法的总体视图，更多地关注电子商务领域的应用场景。与这些调查不同，我们的工作试图确定以前研究中遇到的挑战，并以全面和有条理的方式广泛讨论现有的解决方案。

综述内容

在本文中，我们详细总结了复杂 KBQA 的典型挑战和解决方案。我们首先介绍 KBQA 任务的背景。接下来，我们介绍了复杂 KBQA 的两种主流方法，即基于语义解析（SP-based）的方法和基于信息检索（IR-based）的方法。然后，我们从这两个类别的角度全面回顾了先进的方法。具体来说，我们解释了他们对典型挑战的解决方案。最后，我们总结并讨论了未来研究的一些有希望的方向。

基于复杂问题的问答方法挑战总结

- 1.现有的基于语义解析的方法并不能覆盖各种复杂的查询（如多条推理，约束关系和数值运算），同样，基于信息检索的方法无法回答复杂的查询，因为它们的排序是在没有可追溯推理的小范围实体上执行的。
- 2.在复杂问题中更多的关系和主语表明需要一个巨大的搜索空间来解析问题，这将极大增加计算成本。与此同时，更多的关系和主题将阻止基于信息检索的方法检索所有相关实体去排序
- 3.两种方法都将问题理解视为首要步骤。当问题在语义和句法方面都变得复杂时，需要模型具有强大的自然语言理解和泛化能力。
- 4.对于复杂问题，标记通往答案的真实路径（参见图 1 中的示例）的成本很高。通常，只提供问答对。这表明基于 SP 的方法和基于 IR 的方法必须在没有分别标注正确逻辑形式和推理路径的情况下进行训练。这种微弱的监督信号给这两种方法都带来了困难。

优缺点

基于SP的方法通过生成表达逻辑形式产生更具有可解释性的推理过程，但是他们严重依赖逻辑形式和解析算法的设计。

基于IR的方法适合目前流行的端到端模型，并使基于 IR 的方法更易于训练。然而，推理模型的黑盒风格使得中间推理难以解释。

未来研究方向

- 1.在未来，一个进化的 KBQA 系统对于在线部署后获得持续改进是必不可少的，将用户反馈考虑进去
- 2.这些模型旨在处理分布外问题并获得可解释的推理过程。为 KBQA 设计具有良好可解释性和鲁棒性的方法可能是未来研究的一个具有挑战性但有希望的主题。
- 3.更加通用的知识库