

Case-Based Reasoning for Natural Language Queries over Knowledge Bases

研究动机

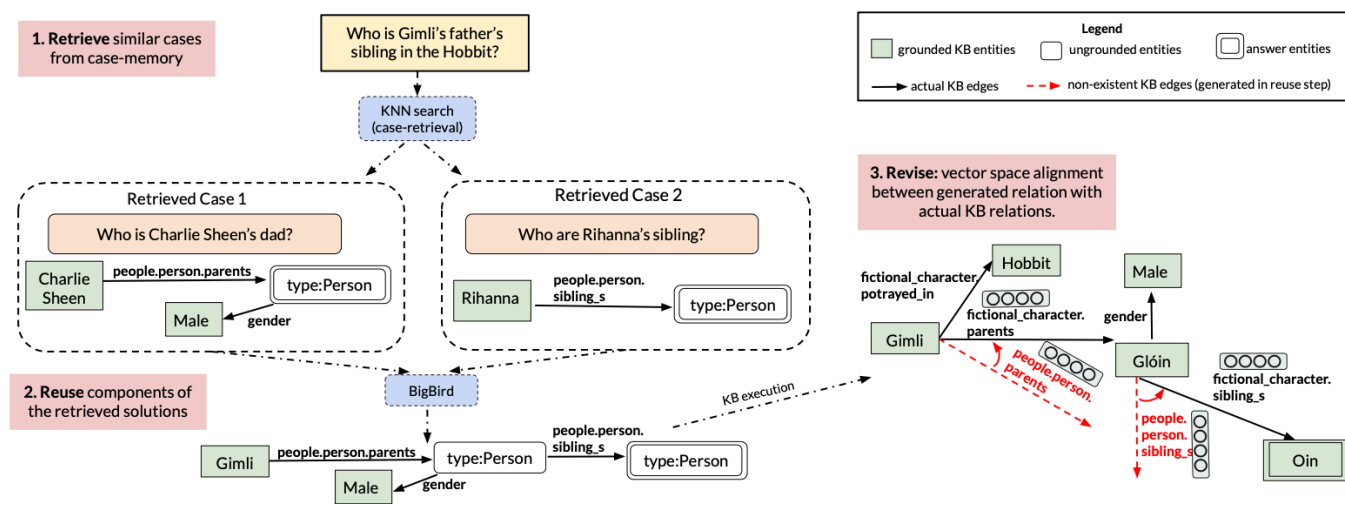
从头开始解决复杂问题通常具有挑战性，但如果我们可以使用他们的解决方案访问其他类似问题，则容易得多。该论文面向有监督的KBQA场景，即每一个query（问题）都被标注了一个对应的logic form（可以看做是一种推理路径），问题的答案可以通过在Knowledge Base上执行问题标注的logic form（推理路径）得到。

研究贡献

我们提出了一种用于KBQA的神经CBR方法，能够生成类似检索问题及其逻辑形式的复杂逻辑形式。（b）由于CBR-KBQA显式地学习重用案例，我们证明了当提供相关案例时，它能够在测试时推广到不可见的关系。（c）我们还展示了CBR-KBQA修改步骤的有效性，该步骤允许通过将生成的输出与查询实体的局部邻域对齐来纠正生成的输出。（d）最后，我们证明了CBR-KBQA在几个KBQA基准上显著优于其他竞争模型。

研究内容

本文旨在为当前query检索出相似的case，然后利用这些相似case的推理路径生成当前query的推理路径，以得到最终的答案。本文模型的大致框架如图所示：



本文提出的模型具体包含三个部分，Retrieve、Reuse以及Revise：

Retrieve： 本文采取使用DPR（Dense Passage Retriever）的方式对query的文本进行编码、检索和训练。其中本文希望得到的检索器能更加注重问题的形式而不是其中包含的具体实体，例如对于问题 谁是小明的哥哥？，本文更希望检索得到 谁是小红的哥哥？ 而不是 谁是小明的爸爸？，因此检索器对文本进行编码时会把实体部分mask掉。

Reuse： 本文采用生成的方式来生成当前query的logic form。具体来说，本文将检索得到的query- logic form对全部拼接起来，然后喂给Seq2Seq模型用自回归的方式生成当前query的logic form。由于拼接起来的输入可能会特别长，本文在这里采用BigBird模型来生成logic form。

Revise： 由于不确定某些生成的关系路径是否存在于Knowledge Base中，本文进一步对生成的logic form进行修正。本文首先用预训练好的TransE模型对Knowledge Base进行编码，然后将生成结果中不存在的关系替换为当前Knowledge Base中与其最相似的关系。

研究结论

据我们所知，我们是第一个为KBQA提出神经化CBR方法的人。我们证明了我们的模型在处理知识库上的复杂问题时是有效的，但我们的工作也有一些局限性。首先，我们的模型依赖于监督逻辑形式（如SPARQL查询）的可用性，在大规模上进行anno-tate可能会很昂贵。未来，我们计划探索直接从问答对中学习的方法（Be-rant等人，2013；Liang等人，2016）。尽管CBR-KBQA是模块化的，并且有几个优点，但我们的模型的检索和重用组件是分别训练的。未来，我们计划探索CBR端到端学习的途径。