

Leveraging Frequent Query Substructures to Generate Formal Queries for Complex Question Answering

研究动机

对于复杂问题，主要有两种查询生成方法。他们可能会受到缺乏训练数据的影响，尤其是对于结构很少出现的长尾问题。此外，上述两种方法都无法处理具有未知查询结构的问题，因为它们无法生成新的查询结构。

研究贡献

- 我们将查询结构的概念形式化，并定义查询结构之间的子结构关系。
- 我们提出了一种新的形式化查询生成方法，该方法首先利用多个神经网络预测给定问题中包含的查询子结构，然后使用组合函数对现有查询结构进行排序。
- 我们合并查询子结构以构建新的查询结构，该结构处理具有未知查询结构的问题。
- 我们在两个KBQA数据集上进行了大量实验，结果表明SubQG显著优于现有的ap方法。此外，SubQG在训练数据有限和实体/关系链接结果有噪声的情况下取得了良好的性能。

研究内容

该研究框架主要是分为线上和线下两部分

线下部分分为三步：

- 1.收集查询结构。对于训练数据中的问题，我们首先发现结构等价的查询，然后提取所有查询结构的集合，用TS表示。
- 2.收集频繁的查询子结构。我们分解每个查询结构 $S_i = (V_i, T_i) \in TS$ 获取所有查询子结构的集合。
- 3.训练查询子结构预测器。我们为每个查询子结构训练一个神经网络 $\pi \in FS$ 公司 π ，预测 Q_y 具有 S_i 的概率 π （即 $S_i \leq [Q_y]$ ），其中 Q_y 表示 y 的形式查询。

线上部分是查询生成过程以自然语言问题 y 作为输入，主要包括四个步骤：

- 1.预测查询子结构。我们首先预测 $S_i \leq [Q_y]$ 对于每个 $S_i \in FS$ 公司 π ，使用在离线步骤中训练的查询子结构预测。
- 2.对现有查询结构进行排序。要为输入问题找到合适的查询结构
- 3.合并查询子结构。考虑到目标查询结构 $[Q_y]$ 可能不适用于TS（即，在结构上等价于 Q_y 的训练数据中没有查询），我们设计了一种方法来合并包含问题的查询子结构，以构建新的查询结构。
- 4.接地和验证。我们利用查询结构排序结果，以及一些现有black box systems的实体/关系链接结果，生成输入问题的可执行正式查询

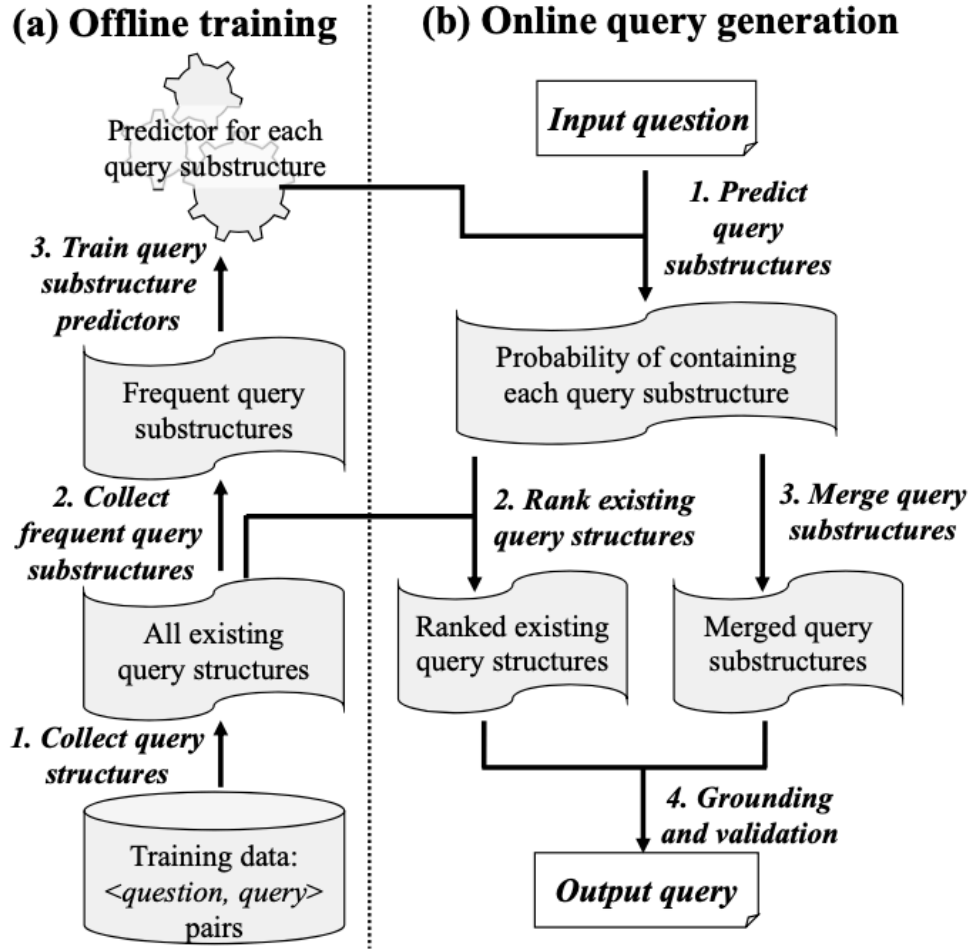


Figure 3: Framework of the proposed approach

研究结论

本文介绍了SubQG，一种基于频繁查询子结构的形式化查询生成方法。子QG首先利用多重神经网络预测问题中包含的查询子结构，然后使用组合函数对现有查询结构进行排序。此外，SubQG合并查询子结构，为训练数据中没有适当查询结构的问题构建新的查询结构。我们的实验表明，SubQG比现有方法取得了更好的结果，尤其是对于复杂问题。在未来的工作中，我们计划添加对其他复杂问题的支持，这些问题的查询需要联合、分组或数字比较。此外，我们对挖掘每个查询子结构的自然语言表达式感兴趣，这可能有助于当前的解析方法。