

Knowledge-enriched, Type-constrained and Grammar-guided Question Generation over Knowledge Bases

研究动机

知识库上的问题生成 (KBQG) 旨在生成关于子图 (即一组三元组) 的自然语言问题。当前基于编码器-解码器的方法仍面临两个主要挑战，尤其是在小子图上：

- (1) 信息不足：由于子图中包含的信息有限，多样性低，流动性差；
- (2) 由于解码器忽略了答案实体的语义，语义漂移。

研究贡献

- 1. 我们用辅助信息扩充源子图以丰富编码器输入，从而提高生成问题的多样性。
- 2. 我们建议在生成的问题中加入单词类型，并使解码器输出以这些类型为条件，从而缓解语义漂移问题。
- 3. 在强化学习框架中，我们设计了一个基于 DPT 的评估器，以鼓励结构一致性，同时不严格执行子序列匹配。
- 4. 我们对两个基准数据集进行了广泛的实验，结果表明，我们的模型大大优于最先进的方法，并且它可以产生更正确、多样化和流畅的问题。

研究框架

我们提出了一种创新的知识丰富、类型约束和语法引导的 KBQG 模型，名为 KTG，以应对上述挑战。在我们的模型中，编码器配备了来自 KB 的辅助信息，而解码器在 QG 期间受到单词类型的约束。具体来说，考虑实体域和描述以及关系层次信息来构建问题上下文，同时结合条件复制机制来根据当前单词类型调整问题语义。此外，还设计了一种具有语法相似性的新奖励函数，旨在通过强化学习提高生成丰富度和句法正确性。

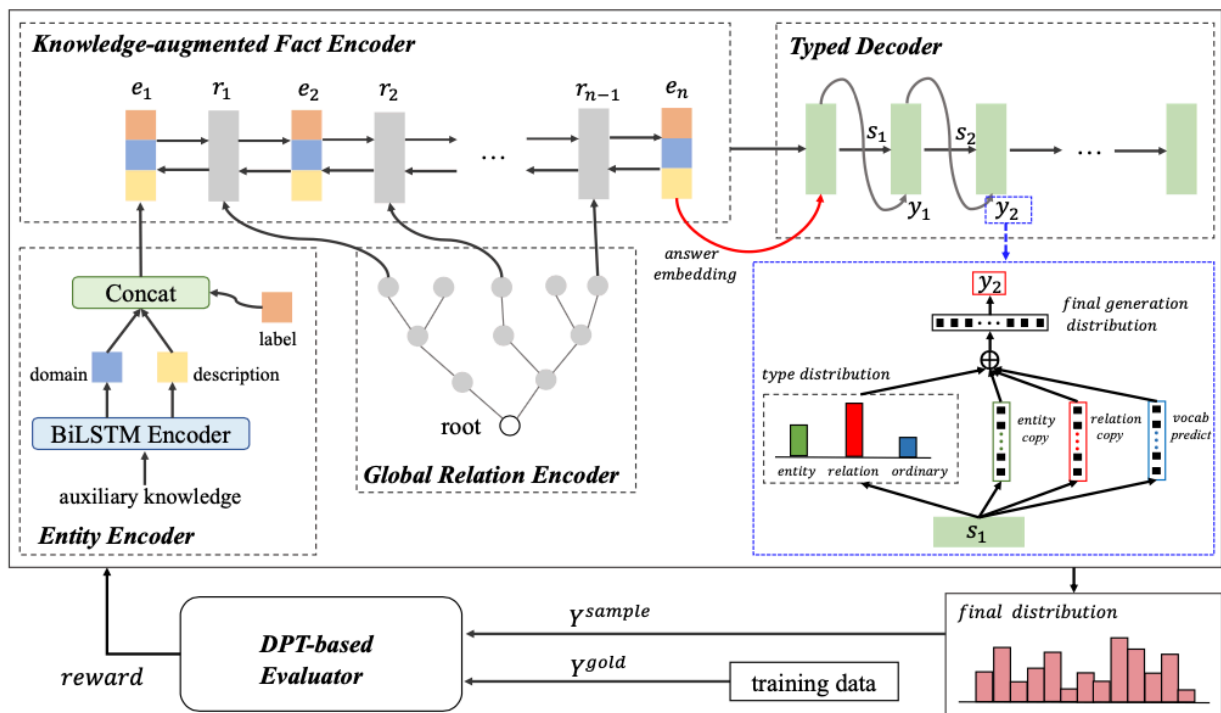


Figure 1: An illustration of our proposed model for question generation over knowledge bases.

研究结论

在本文中，我们解决了两个关键挑战：知识库（KBQG）问题生成任务的源输入不足和语义漂移问题。我们使用辅助知识丰富编码器输入，包括实体描述和谓词域，以提高问题的多样性。我们采用具有条件复制机制的类型化解码器来进一步提高生成问题的语义一致性。我们通过强化学习进一步优化模型性能，并设计了一种基于语法相似性但不基于 n-gram 重叠的新奖励函数。这种奖励确保生成语法和语义上有效的问题，同时允许更多的多样性和流畅性。两个基准数据集的实验结果表明，我们的模型在所有自动和人工评估指标上都比最先进的模型实现了显著改进。将发布源代码以鼓励可重复性和进一步研究