RESCIENCE C

[Re] Improving Multi-hop Question Answering over Knowledge Graphs using Knowledge Base Embeddings

Jishnu Jaykumar P[1, ID] and Ashish Sardana[1, ID]

[1] NVIDIA, Bengaluru, India

1 Introduction

Knowledge is the key to question answering task. Knowledge Graph (KG) is a multi-relational graph consisting of entities as nodes and relations among them as typed edges. KGs can accommodate a wide variety of facts, making them one of the potential candidates for intelligent decision-making. Question Answering over KG (KGQA) task aims to answer natural language queries posed over the KG. Multi-hop KGQA is a trending topic and has gained traction from both academia and industry recently. Multi-hop KGQA task involves reasoning over multiple edges of the KG to arrive at the correct answer.

Earlier works on KGs(e.g. [], [freebase:datadumps], [dbpedia2015], [NELL]) have some element of sparsity, i.e. they do not capture all the facts available in the real world. Recent research on multi-hop KGQA has attempted to reduce this sparsity with the help of relevant external textual resources that are not readily available. On the other side, KG embeddings have emerged as an effective tool to overcome the KG sparsity by predicting missing links in the KG. Although effective, KG embeddings have not been explored for the multi-hop KGQA task. [saxena-etal-2020-improving] fills this gap with the proposed EmbedKGQA method.This work intends to reproduce and perform an ablation (removing relation matching module) as well as extended study on EmbedKGQA(Saxena et al.[2020]). EmbedKGQA claims to be the first of its kind to use KG embeddings for multi-hop KGQA and improves over other state-of-the-art (SOTA) baselines.

2 Scope of reproducibility

According to [saxena-etal-2020-improving], using ComplEx ([ComplEx2016]) KG embeddings significantly improves Hits@1 for multi-hop KGQA task and it has been proved with the help of the results on MetaQA ([metaqa-dataset]) and WebQSP ([webqsp-dataset]) datasets. This reproducibility work tries to test this claim and conducts experiments as mentioned in table:{2,3} of the original paper. Section 4.1 contains the corresponding results which support the claim with some anomalies.
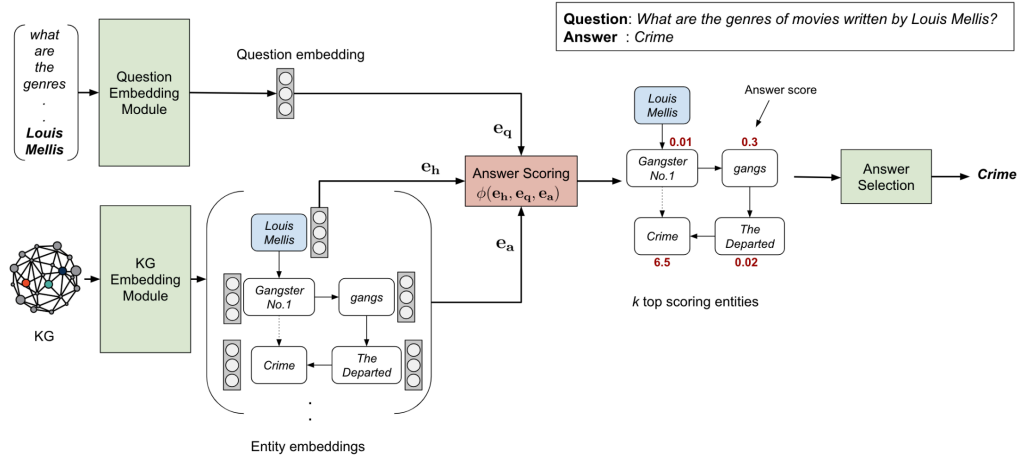
Figure 1. Overview of EmbedKGQA, the proposed method for Multi-hop KGQA.
Image source: [saxena-etal-2020-improving].

3 Methodology

The authors of the original paper have open-sourced the code along with the data and pre-trained ComplEx KG embedding models. We have used the same codebase (commit:5d8fdbd4) and customized it for our purposes. In addition to this, we have added a comprehensive documentation to make it more interpretable. Moreover, a command-line functionality is also added to easily configure various transformers models in the training workflow.

3.1 Model descriptions

As shown in Figure:1, EmbedKGQA has three modules: