

KNOWLEDGE GRAPH-BASED QUESTION ANSWERING WITH MINIMAL SUPERVISION

Abstract

Knowledge Graph Question Answering (KGQA) enables natural language question answering by reasoning over structured Knowledge Graphs (KGs). However, the limited availability of training data annotated with logical forms poses a major challenge to developing effective KGQA systems. This thesis investigates approaches to KGQA under minimal supervision across three settings: zero-shot, few-shot, and low-resource fine-tuning.

In the zero-shot setting, we propose NS-KGQA, a neuro-symbolic approach that translates natural language questions into symbolic question subgraphs using a link prediction model trained solely on the KG, followed by symbolic execution to obtain answers. Building on this foundation, we explore few-shot approaches such as SymKGQA and CodeAlignKGQA that leverage Large Language Models (LLMs) to generate step-by-step logical forms in Knowledge-Oriented Programming Language (KoPL). To improve generation accuracy, we introduce logical-form-specific constraints, question-specific abstract facts for knowledge-aware code generation, and a dynamic self-correction mechanism. Finally, in the low-resource setting, we propose PO-KGQA, a preference-alignment approach that dynamically generates preference data through self-feedback and fine-tunes language models using Low Rank Adaptation (LoRA), enabling them to distinguish correct logical-form sequences from incorrect alternatives.

Overall, this work presents a progression toward KGQA systems that require progressively less supervision while combining the strengths of language models and symbolic reasoning. The proposed approaches aim to improve the accuracy, generalizability, and efficiency of knowledge graph reasoning in data-scarce settings.