Interactive Query Answering on Knowledge Graphs with Soft Entity Constraints

Extended Abstract

Daniel Daza[1], Alberto Bernardi[2], Luca Costabello[2], Christophe Gueret[2], Michael Cochez[3], and Martijn Schut[1]

[1] Amsterdam UMC, Amsterdam, Netherlands
[2] Accenture Labs, Dublin, Ireland
[3] Vrije Universiteit Amsterdam, Netherlands

Abstract. Keywords: Knowledge graphs · Approximate query answering · Interactive systems.

1 Introduction

Knowledge graphs (KGs) serve as structured representations of domain knowledge, supporting reasoning and retrieval tasks. Approximate query answering (QA) methods mitigate KG incompleteness by predicting ranked lists of likely answer entities using embeddings and graph-based models [5, 4, 6, 7, 1, 3, 2]. However, these methods are inherently static: given a query, they return a fixed answer list without allowing further refinement based on user preferences.

For example, given the query "What are the award nominations received by movies starring Leonardo DiCaprio?" a user may want to focus on nominations related to cinematography rather than acting. Existing QA methods lack mechanisms to incorporate such *soft* preferences dynamically.

To address this limitation, we propose an **interactive QA setting** where users provide **soft preferences** to refine answers iteratively. Soft preferences are subsets of entities that either **prioritize** or **de-emphasize** specific types of answers, without enforcing strict logical constraints. This interactive approach enhances the adaptability of KG-based QA, enabling retrieval of answers that are constrained to more specific contexts.

2 Learning to Incorporate Soft Constraints

We extend approximate QA to an interactive setting, where soft preferences iteratively refine the ranked answer set. Given an initial ranking of answers for a query, a user provides a set of preferences as entity-label pairs (e_i, l_i), where e_i is an entity and $l_i \in \{0, 1\}$ denotes whether similar entities should be prioritized (1) or avoided (0). The adjusted scores at iteration $t+1$ are computed as:

$$a^{(t+1)} = f(a^{(t)}, P(t)),$$

where $P(t)$ is the set of preferences provided up to iteration t. The reranking function f adjusts scores dynamically, incorporating the provided soft constraints.

We train a reranking model to refine answer rankings based on user preferences. Given a dataset $\mathcal{D} = \{(q_i, \mathcal{A}_i, P(T_i))\}$, where q_i is a query, $\mathcal{A}_i$ the initial answer set, and $P(T_i)$ the associated soft preferences, the model learns a score adjustment function:

$$a^{(t+1)} = a^{(t)} + f_\theta(e, P(t)),$$

where f_θ is a neural network that modifies entity scores based on preference information. The training objective consists of two ranking losses:

1. **Preference Ranking Loss**: Ensures preferred entities are ranked above non-preferred ones.
2. **Answer Ranking Loss**: Maintains high ranks for correct answers relative to incorrect ones.

3 Benchmarking Interactive QA

To evaluate our approach, we construct benchmarks by augmenting standard KG query datasets with soft preference annotations. Preferences are derived using hierarchical clustering on entity embeddings, partitioning answer sets into semantically coherent subgroups. Our experiments on demonstrate that the reranking model effectively integrates user preferences, achieving significant improvements in preference-aligned ranking metrics. Our results validate the feasibility of interactive KG-based QA and highlight the potential for adaptive, user-driven retrieval.

4 Conclusion

We introduce interactive QA with soft constraints, enabling dynamic refinement of KG query answers. Our framework formalizes this problem, constructs benchmarks for empirical study, and proposes a reranking model that integrates seamlessly with existing QA systems. Results show that soft preferences can be effectively incorporated without retraining base models, opening new avenues for adaptable and user-driven KG-based reasoning.

References

1. Erik Arakelyan, Daniel Daza, Pasquale Minervini, and Michael Cochez. Complex query answering with neural link predictors. In *International Conference on Learning Representations*, 2021.
2. Erik Arakelyan, Pasquale Minervini, Daniel Daza, Michael Cochez, and Isabelle Augenstein. Adapting neural link predictors for data-efficient complex query answering. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, NIPS '23, Red Hook, NY, USA, 2023. Curran Associates Inc.
3. Yushi Bai, Xin Lv, Juanzi Li, and Lei Hou. Answering complex logical queries on knowledge graphs via query computation tree optimization. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett, editors, *International Conference on Machine Learning, ICML 2023, 23-29 July 2023, Honolulu, Hawaii, USA*, volume 202 of *Proceedings of Machine Learning Research*, pages 1472–1491. PMLR, 2023.
4. Daniel Daza and Michael Cochez. Message passing query embedding. In *ICML Workshop - Graph Representation Learning and Beyond*, 2020.
5. William L. Hamilton, Payal Bajaj, Marinka Zitnik, Dan Jurafsky, and Jure Leskovec. Embedding logical queries on knowledge graphs. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, NIPS'18, page 2030–2041, Red Hook, NY, USA, 2018. Curran Associates Inc.
6. Hongyu Ren, Hanjun Dai, Bo Dai, Xinyun Chen, Denny Zhou, Jure Leskovec, and Dale Schuurmans. Smore: Knowledge graph completion and multi-hop reasoning in massive knowledge graphs. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '22, page 1472–1482, New York, NY, USA, 2022. Association for Computing Machinery.
7. Zihao Wang, Yangqiu Song, Ginny Y. Wong, and Simon See. Logical message passing networks with one-hop inference on atomic formulas. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali Rwanda, May 1-5, 2023.* OpenReview.net, 2023.