



Integrating Structured Knowledge into Large Language Models

Evgenii Posashkov, Biljana Vitanova, and Žan Štuhec

Abstract

Large Language Models (LLMs) have demonstrated significant success in natural language processing tasks but still face challenges, one of which is handling complex question answering. This project explores methods to integrate structured knowledge, in form of knowledge graphs (KGs), into LLMs to enhance their performance. We aim to evaluate different techniques for integrating KGs. The ultimate goal is to assess whether the inclusion of structured knowledge can address the existing limitations of LLMs, and improve their reasoning capabilities.

Keywords

Large Language Models, Knowledge Graphs, Question Answering, Reasoning

Advisors: Slavko Žitnik

Introduction

Despite advances in text generation by LLMs, they struggle with complex reasoning and fact recall. One approach to overcoming these limitations is integrating Knowledge Graphs (KGs), which provide structured representations of entities and relationships.

Our project explores methods to incorporate KGs into LLMs by pairing questions with relevant KGs, bridging the gap between unstructured text and structured knowledge, enabling LLMs to access and utilize detailed information for advanced reasoning and accurate conclusions.

By using both textual and structural elements of KGs, we aim to improve the accuracy of LLMs in complex question-answering scenarios.

Related Work

Recent research has explored multiple ways to integrate KGs into LLMs to better reasoning and improve task performance. Tang et al. (2024) introduced GraphGPT [1], which pairs LLMs with graph structures, aligning closely with our goal of improving question answering through KG integration. Kau et al. (2024) examined how KGs can mitigate issues like hallucinations in LLMs [2], which is relevant to our focus on improving factual accuracy. Yao et al. (2019) developed KG-BERT for knowledge graph completion [3], inspiring our use of graph embeddings in LLMs for better performance in complex reasoning tasks. Yasunaga et al. (2022) proposed

DRAGON, a model that fuses text and KG during pretraining [4], which relates to our aim of combining textual and structural elements of KGs to enhance LLM capabilities. Kang et al. (2023) presented SURGE for knowledge-grounded dialogue [5], which aligns with our goal of ensuring that KGs improve LLMs factual accuracy in more complex tasks. And in the end Wang et al. (2017) surveyed KG embedding techniques [6], providing the basis for our exploration of graph embeddings like TransE and node2vec to better LLM performance. These studies together inform our methodology for integrating KGs into LLMs to improve their reasoning and accuracy in complex question answering.

Representation and Integration of KGs

KGs and LLMs can be integrated in two ways: KG-enhanced LLMs, where knowledge graphs improve the responses of large language models, and LLM-enhanced KGs, where language models enhance the representation of knowledge graphs [7]. In our project, we will focus on the first approach—KG-enhanced LLMs, or integration of KGs in LLMs.

KGs can be integrated into LLMs at different stages on model's development. Based on when the integration occurs, we can distinguish between two types of KG enhancement.

1. During-training enhancement: KGs are integrated into the model's training process, modifying the model's architecture to recognize this structured information.

2. Post-training enhancement: After the model has been trained, KGs are used to refine the performance by providing additional information through the prompt itself.

Additionally, Knowledge Graphs can be classified based on the type of knowledge they represent:

1. Encyclopedic KGs – These KGs store broad real-world knowledge, providing general information. Examples include Wikidata, NELL, DBpedia.
2. Commonsense KGs – which capture logical relationships from everyday life. Examples are: ConceptNet, ATOMIC.
3. Domain-Specific KGs – Focus on specialized fields like medicine or finance. One example is UMSL which is used in the medical field.
4. Multimodal KGs – Combine structured data with multiple sources, such as images and text. IMGpedia is an example of such.

Methodology

In the following section, we will introduce our primary plan on how we aim to approach this project. There are four key questions we need to answer: Which integration techniques will be applied, which LLMs will be used as baselines, how will the evaluation dataset be created, and how will the evaluation be performed.

Knowledge Graph Integration Techniques

In our implementation, our main focus will be on during-training enhancement, as we expect it to yield better performance. To achieve this, we will experiment with different graph embeddings such as TransE, LapEig, node2vec, and DeepWalk, and analyze their performance. We will explore integrating attention-based fusion techniques, where the model uses its attention mechanism to focus on the most relevant parts of the knowledge graph, allowing it to dynamically combine structured knowledge with textual data for improved predictions

Baseline LLMs

As baselines we will use already pretrained models like GPT, BERT, T-5 available online.

Evaluation Dataset

For the final work, we plan to use Slovenian linguistic data from the "Slovenian Digital Dictionary Database". As we do not have access to this data yet, our starting point will be datasets that contain complex questions [8],[9], which require different levels of reasoning. For example:

1. one-hoop: What movies are about [George Pal]? -

2. two-hoop: What movies are about [a director who won an Oscar]? (two-hop)
3. three-hoop: Which person wrote the films directed by the director of [Four Brothers]?

Questions will be paired with a relevant knowledge graphs extracted from sources that provide structured data, such as DBpedia, Wikidata or similar databases. Additionally, we will explore how to pair multiple knowledge graphs and what makes a well-structured knowledge graph.

Performance Evaluation

Performance evaluation will be done from several aspects. First, we'll check if the answers are correct and relevant. Then, we'll assess how well the model handles questions of different complexity, especially those requiring multi-step reasoning. We'll also measure response time to see how quickly the model generates answers. Finally, we'll evaluate how well the model uses structured data from knowledge graphs to improve its answers.

Discussion

We acknowledge that some aspects of the methodology may not be fully accurate at this stage, as our focus has been primarily on theoretical understanding. Through implementation, we expect to gain further insights and be able to provide a more detailed and refined explanation.

References

- [1] Jiabin Tang, Yuhao Yang, Wei Wei, Lei Shi, Lixin Su, Suqi Cheng, Dawei Yin, and Chao Huang. Graphgpt: Graph instruction tuning for large language models. In *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 491–500, 2024.
- [2] Amanda Kau, Xuzeng He, Aishwarya Nambissan, Aland Astudillo, Hui Yin, and Amir Aryani. Combining knowledge graphs and large language models. *arXiv preprint arXiv:2407.06564*, 2024.
- [3] Liang Yao, Chengsheng Mao, and Yuan Luo. Kg-bert: Bert for knowledge graph completion. *arXiv preprint arXiv:1909.03193*, 2019.
- [4] Michihiro Yasunaga, Antoine Bosselut, Hongyu Ren, Xikun Zhang, Christopher D Manning, Percy S Liang, and Jure Leskovec. Deep bidirectional language-knowledge graph pretraining. *Advances in Neural Information Processing Systems*, 35:37309–37323, 2022.
- [5] Minki Kang, Jin Myung Kwak, Jinheon Baek, and Sung Ju Hwang. Knowledge graph-augmented language models for knowledge-grounded dialogue generation. *arXiv preprint arXiv:2305.18846*, 2023.

- [6] Quan Wang, Zhendong Mao, Bin Wang, and Li Guo. Knowledge graph embedding: A survey of approaches and applications. *IEEE transactions on knowledge and data engineering*, 29(12):2724–2743, 2017.
- [7] S. Pan, L. Luo, Y. Wang, C. Chen, J. Wang, and X. Wu. Unifying large language models and knowledge graphs: A roadmap. *IEEE Transactions on Knowledge and Data Engineering*, 36(7):3580–3599, 2024.
- [8] Alon Talmor and Jonathan Berant. The web as a knowledge-base for answering complex questions. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 641–651, 2018.
- [9] Yuyu Zhang, Hanjun Dai, Zornitsa Kozareva, Alexander J Smola, and Le Song. Variational reasoning for question answering with knowledge graph. In *AAAI*, 2018.