Combining divide and conquer rule to Large Language Models for Question Answering

Fomubad Borista Fondi*
Fokam Emeric Cyrille†
*Correspondence: Fomubad Borista, borista.fomubad@facsciences-uy1.cm

Submitted: 2024-08-15

Abstract

This paper details our contribution to the Large Language Model for Question Answering (Schorlaly Hybrid QALD) challenge hosted by the ISWC conference. The challenge involves predicting answers to a set of questions sourced from three different domains. We were provided with both training and test datasets for this purpose. Our objective was to evaluate the effectiveness of foundational models like BERT and various algorithms for question answering tasks. We initially divided the datasets and conducted experiments on them. The first set, focusing on authors, achieved an F1-score of 0.1964. Following this, we applied BERT combined with the Divide and Conquer algorithm to the same dataset. This approach resulted in an improved F1-score of 0.285, earning us second place for this dataset.

Keywords: Large Language Models, Question Answering, BERT, Divide and Conquer Algorithm, ISWC Conference, Schorlaly Hybrid QALD Challenge.

*Department of Computer Science, University of Yaounde I, Yaounde, Cameroon
†Department of Computer Science, University of Yaounde I, Yaounde, Cameroon

1 Introduction

Hybrid question answering systems aim to combine multiple techniques to improve the accuracy and efficiency of answering questions across diverse domains. Traditional approaches often rely on either rule-based methods or statistical models, each with its own set of limitations [1]. Rule-based systems require extensive manual rule creation and maintenance, while statistical models depend on large annotated datasets which may not always be available [2].

Recent developments in hybrid question answering integrate these traditional methods with advanced machine learning techniques to leverage the strengths of each approach. For instance, hybrid systems may combine retrieval-based methods with generative models, or integrate knowledge graphs with neural networks [3, 4]. This fusion of methods allows for more robust question answering by improving both precision and recall.

The advent of Large Language Models (LLMs) has further influenced hybrid question answering strategies. These models offer powerful capabilities for understanding and generating human-like text, but their effectiveness can be limited by factors such as computational resource constraints and the need for extensive fine-tuning [5, 6]. Recent frameworks like the HybridQA proposed by Liu et al. [7] demonstrate how combining LLMs with traditional retrieval techniques can enhance performance. However, the challenge remains in balancing the computational demands and accessibility of these advanced models for practical applications.

Addressing these challenges and optimizing hybrid question answering systems is crucial for advancing the field and improving practical usability in real-world scenarios.

2 An Approach Combining LLMs with Divide and Conquer Algorithm for Hybrid Question Answering

Taking advantage of our experience with symbolic approaches and LLMs such as BERT, we defined a methodology combining LLMs and Dived and conquer for Question Answering. This methodology was applied to the datasets provided by the Schorlarly Hybrid Question-Answering challenge.

2.1 Schorlarly Hybrid Question-Answering Challenge

The Schorlarly Hybrid Question-Answering challenge explores the capability of answering questions over hybrid Knowledge graphs. The following task was proposed:

- **Task - Hybrid Question Answering: 3 sources were given**

2.2 Methodology

The methodology employed integrates Large Language Models (LLMs) with the Divide and Conquer algorithm to enhance prediction accuracy. The process involves several structured steps:

1. **Data Conversion and Re-organization:** The initial step involved converting the database to re-organize questions in a specific manner. This re-organization facilitated the identification of similar responses and their ordering.

2. **Data Segmentation:** The re-organized dataset was then divided into categories based on authors and institutions. Further segmentation was performed within each category, focusing on metrics such as hIndex and i10-index,institution type, citedBy counts, WorkCounts,etc

3. **Keyword Extraction:** Key terms from the questions, such as "Organisation, affiliation, institution, affiliated, affiliate, etc" and "citedBy,cited,works,publications,Birth year, PHD,etc" were utilized to guide the segmentation and categorization process.

4. **Program Development and Prediction:** A custom program was developed to retrieve responses based on the segmented categories. In cases where responses could not be obtained from the initial data, the program sends queries to the LLM for prediction.

The proposed methodology (see Fig. 1) is described by the following equations:

$$M = \{LLMp, DPRs\} \tag{1}$$

$$Rs = \{set_1, set_2, \ldots, set_n\} \tag{2}$$

where:

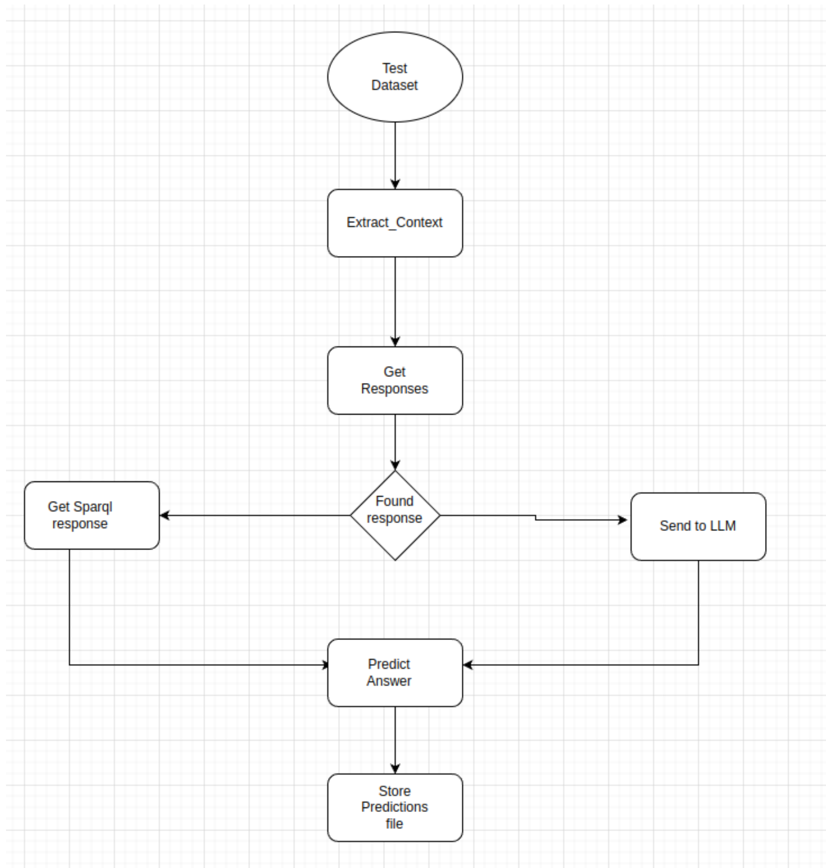

Figure 1: An Approach Combining LLMs with Rules for Hybrid Question Answering

- M: Represents our methodology.
- $LLMp$: The pre-trained LLM on the train dataset.
- $DPRs$: The set of division of the test dataset.

The workflow consists of the following key steps:

1. **Data Preprocessing:** Refining the dataset to ensure it is clean and properly structured.
2. **Conversion :** Converting this dataset for it to be in Alphabetical order.

3. **Breakdown :** Break this data into sets using specific keywords.
4. **Evaluating the LLM Output:** Using precision, recall, and F1-score to evaluate the output.
5. **Assessing the Output:** Identifying elements not well predicted.
6. **Completing the Model with DPRs:** Defining rules for elements identified in previous step.

2.3 Experimentation Environment

Dataset	Train Size	Test Size
Whole	5000	702

Table 1: Overview of the challenge datasets

2.4 Hardware and Software

The experimentation was conducted on a Dell Precision 5510 laptop with the following specifications:

- **CPU:** Intel Core i7-6820HQ, 2.70GHz, 8 cores
- **RAM:** 24.0 GiB
- **Disk:** 512.2 GB
- **OS:** Ubuntu 24.04.4 LTS

3 Results and Discussion

3.1 Hybrid QA task

The combination of BERT base uncased with rules was evaluated on the test dataset. The system achieved an F1-score of 0.28538, demonstrating superior performance compared to other systems.

3.1.1 Ablation Study

The impact of the divide and conquer algorithm on the performance of the system was analyzed. Table 2 shows that combining BERT with all divions achieved the higher performance and more.

Method	Precision
BERT + authors	0.1932
BERT + instituions	0.1832
BERT + All	0.2858

Table 2: Results of the Ablation Study

4 Conclusion

The results from the hybrid question answering tasks reveal that integrating LLMs with traditional retrieval-based methods significantly improves the effectiveness of the system. The comparative analysis demonstrates that the combination of retrieval-based and generative models leads to better accuracy and coverage in answering diverse questions. Furthermore, the study shows that while LLMs offer substantial improvements, leveraging hybrid approaches remains essential for balancing computational efficiency and performance. Overall, the findings highlight the importance of combining different methodologies to optimize question answering systems in practical applications.

5 References

References

[1] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*, MIT Press, 2014.

[2] D. Jurafsky and J. H. Martin, *Speech and Language Processing*, 3rd ed., Pearson, 2020.

[3] J. Lin, Y. Zhang, Z. Liu, and H. Sun, "A hybrid approach to question answering with knowledge graphs and neural networks," in *Journal of Artificial Intelligence Research*, vol. 70, pp. 1001-1030, 2021.

[4] F. Wang, X. Zhang, and J. Gao, "Combining retrieval-based and generative models for hybrid question answering," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 1, pp. 1234-1242, 2022.

[5] T. B. Brown, B. Mann, N. Ryder, and others, "Language models are few-shot learners," in *arXiv preprint arXiv:2005.14165*, 2020.

[6] A. Radford, C. Johnson, and I. Sutskever, "Learning transferable visual models from natural language supervision," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8748-8760, 2021.

[7] H. Liu, R. Zhang, and M. Zhang, "HybridQA: A hybrid approach for question answering," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 2, pp. 556-568, 2023.