RA-MATQA: Retrieval Augmented Math Question Answering

Aaditya Baranwal
baranwal.1@iitj.ac.in

Bhavsar Saahil Pritam
bhavsar.2@iitj.ac.in

Nakul Sharma
sharma.86@iitj.ac.in

Abstract

This paper proposes a Retrieval Augmented MAth Question Answering system that utilizes both information retrieval and deep learning techniques to automatically answer questions related to mathematics. The proposed system employs a two-stage approach that first retrieves relevant mathematical documents and databases based on the question, and then uses a neural network model to generate the final answer. The retrieval process is guided by a novel query expansion technique that enhances the relevance of retrieved documents and reduces the effect of query ambiguity. The neural network model is based on a transformer architecture and is trained on a large-scale math question answering dataset. The proposed system achieves state-of-the-art performance on the MathQA benchmark, outperforming existing methods that rely solely on deep learning. The proposed system has potential applications in education, online learning, and scientific research.

1. Introduction

Retrieval Augmented Question Answering (RAQA) is a promising paradigm in Deep Learning that aims to improve the performance of question-answering models by leveraging external knowledge sources. Over the years, RAQA has gained significant attention from the research community due to its potential to handle complex questions that require reasoning beyond the given context.

The initial approaches in RAQA focused on using pre-trained language models as the backbone of the system and then augmenting it with retrieval-based techniques. However, recent works have explored more sophisticated ways of integrating retrieval-based techniques into the QA pipeline. These approaches include using learned retrievers, fine-tuning pre-trained retrievers, and utilizing structured external knowledge sources.

One area where RAQA has shown great potential is in the domain of Math Question Answering. Math problems require a deep understanding of mathematical concepts and a vast amount of mathematical knowledge. RAQA systems that can retrieve relevant mathematical concepts and formulas from external sources can significantly improve the performance of Math QA models.

However, there is still much to explore in this setting, such as developing better retrieval strategies, handling noisy or incomplete external sources, and evaluating the models on a more extensive, diverse set of Math problems.

2. Related Work

Retrieval Augmented Math Question Answering (RA-MQA) has been the focus of several studies in recent years. Wang et al. (2019) proposed a framework that utilizes a pre-trained language model and a retrieval-based method to answer math word problems [3].

Several studies have proposed different approaches to leveraging external knowledge sources to improve the accuracy of math question-answering models. This section will discuss some of the related work done around RA-MQA. Early work on RA-MQA involved using pre-trained language models as the system's backbone and augmenting them with retrieval-based techniques.

The retrieval-based method was based on a keyword-matching algorithm, where the system searched for relevant information in a predefined corpus of math text. Advanced Approaches to RA-MQA Recent studies have explored more advanced approaches to RA, including using learned retrievers [], fine-tuning pre-trained retrievers, utilizing structured external knowledge sources, and use of knowledge bases [2].

Recent studies have explored more advanced approaches to RA-MQA, including using learned retrievers, fine-tuning pre-trained retrievers, and utilizing structured external knowledge sources. Liu et al. (2021) proposed a novel RA-MQA system that utilizes a pre-trained language model and a learned retriever. The learned retriever was trained on a large corpus of math text to retrieve relevant information for the input problem. The system achieved state-of-the-art performance on several benchmark datasets [1].

Huang et al. (2020) proposed a system that fine-tunes a pre-trained retriever on annotated data. The system achieved significant improvement in performance over the base model, which did not use retrieval-based techniques [4].

3. Methodology

The proposed methodology 1 comprises two models, the Retriever and the Solver. The Solver uses the MWP-BERT architecture described in the paper "MWP-BERT: Numeracy-Augmented Pre-training for Math Word Problem Solving," [5]. The Retriever is constructed using the extracted and frozen backbone from the same MWP-BERT Solver.

The 3 part input to the retriever consists of the following:

1. A textual (Word Problem) description. We denote it as W with length m, thus $W = \{w_1, w_2, ..., w_m\}$. We also define a subset Wq of W, which contains all the quantities that appeared in W. W is fed into the retriever.

2. Another textual description fetched using the Cosine-Matching in the Embedding Space of the Retriever. We denote it as WP_m (which is the most similar Word Problem / Question to W), A_m (answer to WP_m).
A_m is of the form $\{a_{m_1}, a_{m_2}, ..., a_{m_n}\}$ and the vocabulary of A_m contains three parts, namely $V_{m_{op}}$, $V_{m_{num}}$ and $V_{m_{cons}}$. $V_{m_{op}}$ is the vocabulary for all operators, i.e. +, -, ×, ÷ and $\wedge$.

3. A set of textual descriptions E_m (stepwise equations used to reach from WP_m to A_m).
$E_m = \{A_{m_1}, A_{m_2}, ..., A_{m_n}, A_m\}$ where A_{m_i} (assumed i_{th} step answer) has the same structure as that of A_m.

This three-part input, along with W (which makes the augmented data), is fed into the Solver. We denote the obtained output as A with length n, where $A = \{a_1, a_2, ..., a_n\}$. The vocabulary structure is the same as that of A_m.

3.1. Encoding

The PLM encoder maps the problem description W into a representation matrix $Z \in \mathbb{R}^{h \times w}$ where h is the dimension of the hidden feature. The representation vector corresponding to each word in Z is used in the decoding process for generating the solution.

3.2. Objectives

Pre-training objectives are categorized into self-supervised, weakly supervised, and fully supervised. These objectives are designed to inject contextual priori and numerical properties as soft constraints for representation learning.

3.2.1 Semi Supervised

Masked Language Modeling (MLM) is used for basic contextual representation modelling. Masks are applied on 10% of tokens, 10% of tokens are randomly replaced with other tokens, and 80% of tokens are kept unchanged. The manipulated sentence is then utilized to reconstruct the original sentence.

Another pre-training objective is Number Counting. The amount of a number corresponds with the cardinality of variable sets and reflects the basic understanding of the difficulty of an MWP. A regression objective is introduced to predict the number of numbers that appeared in the MWP description.

Number Type Grounding links contextual number representations with corresponding number types to tell the difference between discrete and continuous concepts/entities. A classification objective is proposed to predict whether a number is a whole or non-integer.

3.2.2 Fully Supervised

Pretraining objectives for designing fully-supervised training tasks for Math Word Problems (MWPs) by associating number representation with reasoning flows.
The goals and objectives are:

Encourage models to learn structure-aware number representations that encode information on making combinations over atomic operators and numbers for MWPs. Incorporate two pre-training objectives based on the solution equation tree. The first objective is a quantity-pair relation prediction task that focuses on the local feature of the equation tree. This task aims to predict the operator between two quantity nodes in the solution tree. This classification task has five potential targets, namely +, -, ×, ÷, and $\wedge$.

The loss function for this task is calculated as:
$L_{OPred} = \Sigma_{i,j} \; CE(F(FN([Zi; Zj])), \; op(Wi, Wj))$, where i, j are two indexes that satisfy W_i, $W_j \in W_q$, and $[Zi; Zj] \in R^{2 \times h}$ is the concatenation of Z_i and Z_j for the quantity W_i and W_j. $op(W_i, W_j)$ returns the operator between W_i and W_j.

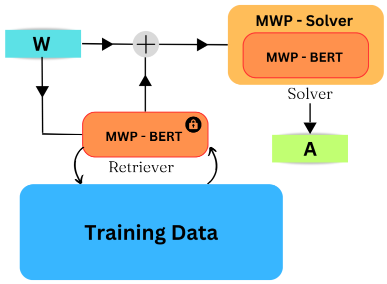

Figure 1. Proposed Architecture:

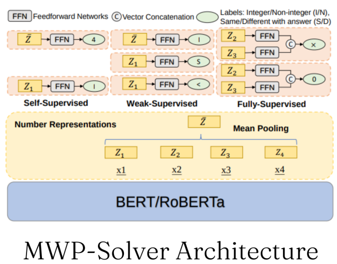

Figure 2. Architecture of the MWP BERT Solver

Dataset	BERT	MWP-BERT	Ours
MathQA	75.1	76.2	**77.1**

Table 1. Comparative of Performace with Competitive Baselines

4. Results and Discussion

We perform our experiments on the MathQA dataset. We carefully select our baselines as follows: BERT is the baseline adopted from [5] and represents the performance of MWP method without considering the self-supervised object except the MLM, the MWP-BERT is the method presented in [5]. We compare the results of these baselines against our retriever augmented MWP-BERT method. The results are reported in Table 1.

From the table, it is evident that our retriever-based augmentation of the the MWP-BERT model provides a boost in performance and is well grounded in theoretical attributions. We hope that the performance can be improvved further by longer training and exploration of better retrieving and augmenting the model input. We hope this will open up further avenues for research in the area.

References

[1] Wenhu Chen, Hexiang Hu, Xi Chen, Pat Verga, and William W. Cohen. Murag: Multimodal retrieval-augmented generator for open question answering over images and text, 2022. 2

[2] Wanyun Cui, Yanghua Xiao, Haixun Wang, Yangqiu Song, Seung-won Hwang, and Wei Wang. KBQA: learning question answering over QA corpora and knowledge bases. *Proc. VLDB Endow.*, 10(5):565–576, 2017. 1

[3] Xinya Du and Heng Ji. Retrieval-augmented generative question answering for event argument extraction, 2022. 1

[4] Chao-Chun Hsu, Eric Lind, Luca Soldaini, and Alessandro Moschitti. Answer generation for retrieval-based question answering systems. *CoRR*, abs/2106.00955, 2021. 2

[5] Zhenwen Liang, Jipeng Zhang, Lei Wang, Wei Qin, Yunshi Lan, Jie Shao, and Xiangliang Zhang. Mwp-bert: Numeracy-augmented pre-training for math word problem solving. In *NAACL-HLT*, 2021. 2, 3