

QA System as an Answer Retrieval Problem

Anonymous ACL submission

Abstract

We plan to implement a Question Answering (QA) system for the financial industry. We plan to implement this using Transformer based models by restricting the scope to just the financial domain. We expect the input data to be a free-form text. We intend to explore new architectures involving pretrained models and fine tune them to retrieve the best answer from a large set of answers for a given question.

This research is not a first of its kind. We surveyed the literature and we found some very interesting related papers. Maia et.al [1] and Bithiah Yuan [4] used pairwise learning. While Maia et.al. did a comparative study that leveraged question context on pairwise learning, Bithiah Yuan focused on financial non-factoid answer selection and retrieved a set of passage-level texts and selected the most relevant as the answer. Zhiyu Chen et.al. [3] performed numerical reasoning over financial data. Suman et. al. [5] provided a simplified approach and used SQuAD dataset to demonstrate their approach. Zhuang Liu et. al. [2] presented a domain specific language model pre-trained on large-scale financial corpora that enabled the capture of language knowledge and semantic information.

Our goal is to deliver a product that will help financial advisors to answer questions based on financial reports and/or disclosures. Often, there are inconsistencies and ambiguities in human interpretation. The outcome of our project will supplement FAQs automation, Chatbox, and dialog systems pertinent to the financial domain.

1 Introduction

Add introduction section here.

2 Related Work

Related work goes here.

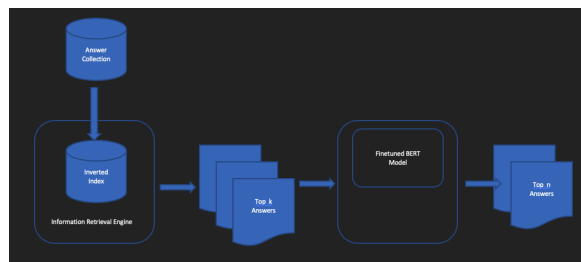


Figure 1: Baseline Model.

3 Our work

We will try 3 approaches (1) Baseline Approach, (2) Two tower approach, and (3) Sentence Embedding.

4 Baseline Approach

Overview of our baseline QA pipeline. The inverted index retriever first returns the top k candidate answers. We will use the pretrained BERT mode with the financial corpus and by finetuning the pretrained BERT model to the target FiQA dataset. The output of the BERT model will pass the output to a softmax layer to assign a probability for each of the selected k answers and finally we will output the top n images within the selected k images.

5 Two Tower Approach

Papers that have been or will be submitted to other meetings or publications must indicate this at submission time in the START submission form, and must be withdrawn from the other venues if accepted by ACL 2020. Authors of papers accepted for presentation at ACL 2020 must notify the program chairs by the camera-ready deadline as to whether the paper will be presented. We will not accept for publication or presentation the papers that overlap significantly in content or results with papers that will be (or have been) published elsewhere.

6 Approach using Sentence Embedding

Papers that have been or will be submitted to other meetings or publications must indicate this at submission time in the START submission form, and must be withdrawn from the other venues if accepted by ACL 2020. Authors of papers accepted for presentation at ACL 2020 must notify the program chairs by the camera-ready deadline as to whether the paper will be presented. We will not accept for publication or presentation the papers that overlap significantly in content or results with papers that will be (or have been) published elsewhere.

7 Metrics

Papers that have been or will be submitted to other meetings or publications must indicate this at submission time in the START submission form, and must be withdrawn from the other venues if accepted by ACL 2020. Authors of papers accepted for presentation at ACL 2020 must notify the program chairs by the camera-ready deadline as to whether the paper will be presented. We will not accept for publication or presentation the papers that overlap significantly in content or results with papers that will be (or have been) published elsewhere.

8 Results

Papers that have been or will be submitted to other meetings or publications must indicate this at submission time in the START submission form, and must be withdrawn from the other venues if accepted by ACL 2020. Authors of papers accepted for presentation at ACL 2020 must notify the program chairs by the camera-ready deadline as to whether the paper will be presented. We will not accept for publication or presentation the papers that overlap significantly in content or results with papers that will be (or have been) published elsewhere.

9 conclusion

Papers that have been or will be submitted to other meetings or publications must indicate this at submission time in the START submission form, and must be withdrawn from the other venues if accepted by ACL 2020. Authors of papers accepted for presentation at ACL 2020 must notify the program chairs by the camera-ready deadline as to

whether the paper will be presented. We will not accept for publication or presentation the papers that overlap significantly in content or results with papers that will be (or have been) published elsewhere.

Acknowledgments

The acknowledgments should go immediately before the references. Do not number the acknowledgments section. Do not include this section when submitting your paper for review.

A Appendices

Appendices are material that can be read, and include lemmas, formulas, proofs, and tables that are not critical to the reading and understanding of the paper. Appendices should be **uploaded as supplementary material** when submitting the paper for review. Upon acceptance, the appendices come after the references, as shown here.