

DevRev

Team 31

Team Members

Team 31

- Anand Hegde
 - Nirmitt Arora
 - Chidaksh Ravaru
 - Agrim Jain
 - Shubh Agarwal
 - Ninad Khati
 - Keshav Naram
 - Saloni Singh
 - Devesh Kumar
 - Rajat Lavekar
-

How do we answer the
given questions
correctly and ASAP?

Sub Tasks of Problem

—

Sub Tasks

Solving individual sub problems
to get better performance!

- Synthetic Data Generation
 - Look for already answered questions in Train and Test Data
 - Find the paragraph of previously answered questions
 - Answer the new question from the paragraph
-

Synthetic Data Generation

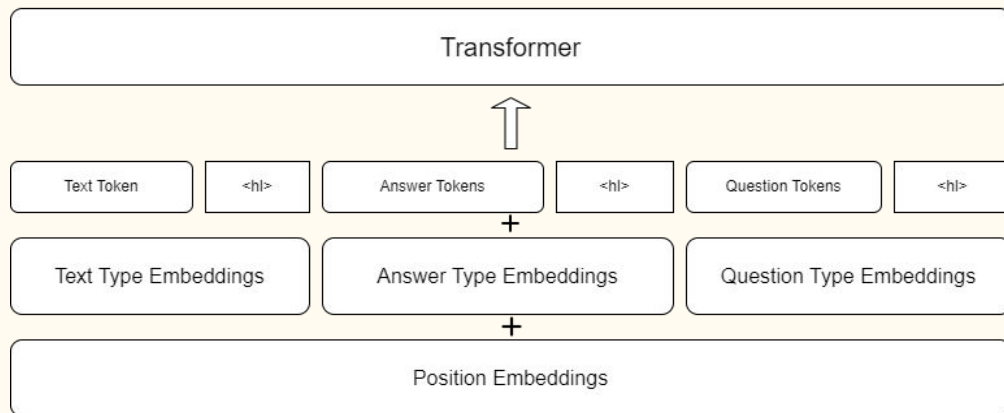


- We used a T5-model to generate synthetic data of 4186 new questions.
- The model was trained to extract spans from the sentences which are considered as answers.

- Now a decoder model generates questions to generated answers.
- Generates better quality data than other methods.
- Average time taken to generate one question is 6.659 seconds.

Question Generation

Pretrained t5-base-qq-hl model was used, its working is described below.

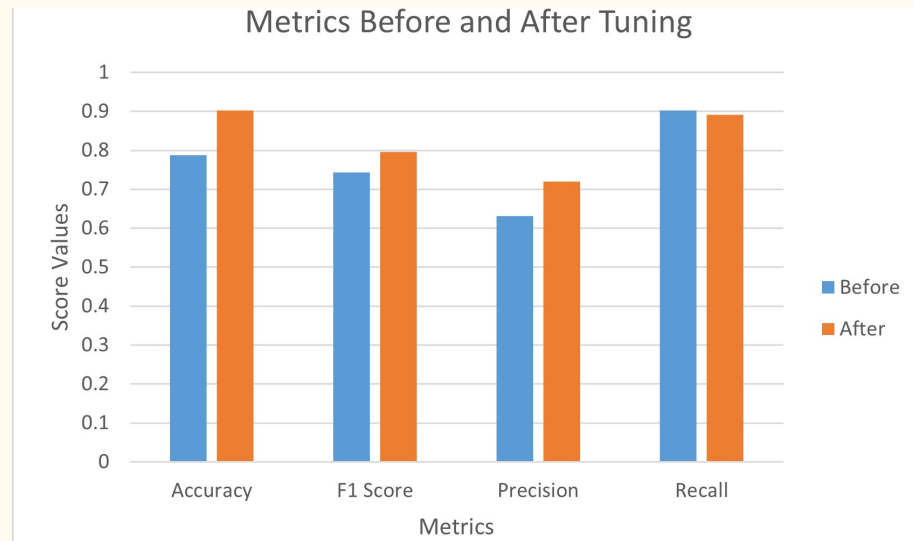


Previously Answered Questions

—

We look for the repeated
questions in the Train
and Test Data...

- MINI-LM model like architecture used to find sentence embeddings.
- Saved training questions' embeddings with the answer and paragraph in file.
- Add new testing questions as we answer them.
- Fast and Efficient.



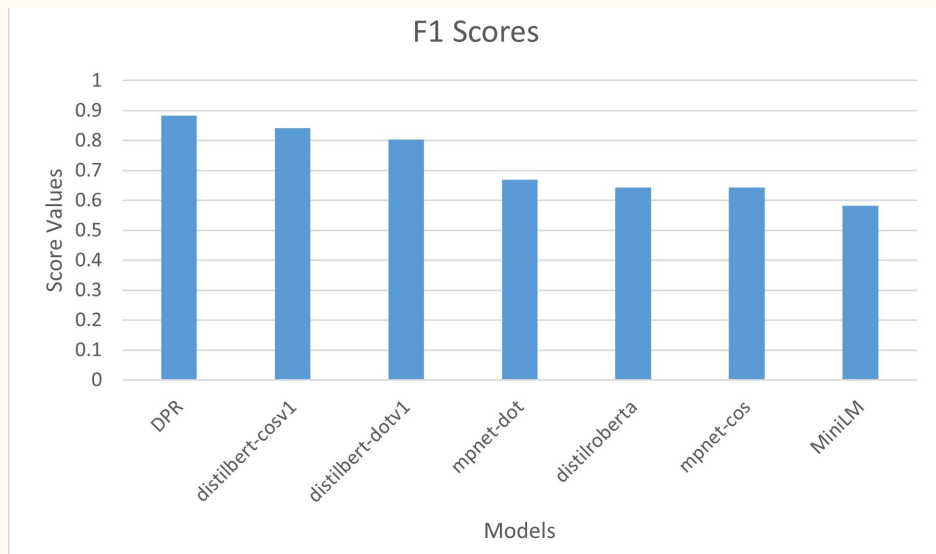
Avg Model Runtime per question: 11.5 milliseconds

Finding Paragraph

—

We look for the
paragraph of the given
theme which can answer
the question...

- We tried different passage retrieval models like DPR, MultiQA-distilbert-cosv1, etc.
- DPR showed us Promising results (f1 score of 0.88) but the inference time for each query was very high.
- So we used MultiQA-distilbert-cosv1 as our retriever (f1 score of 0.84)



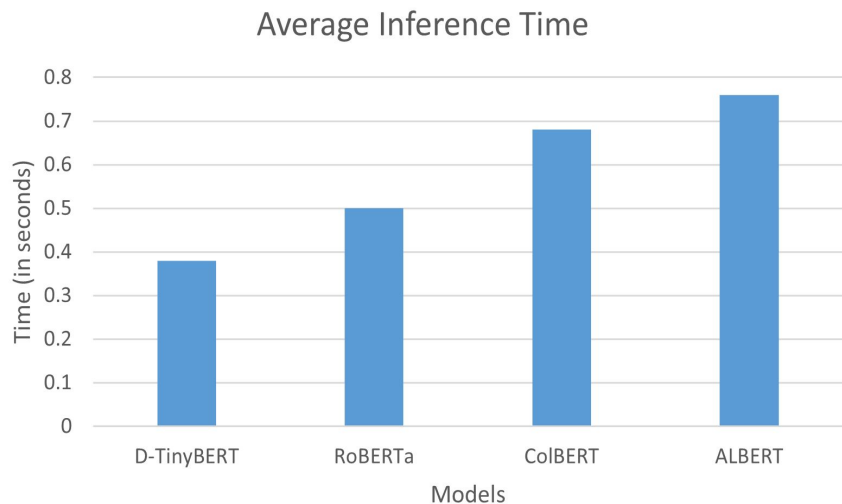
Avg. paragraph finding time per question: 0.115 seconds

Answering the Questions



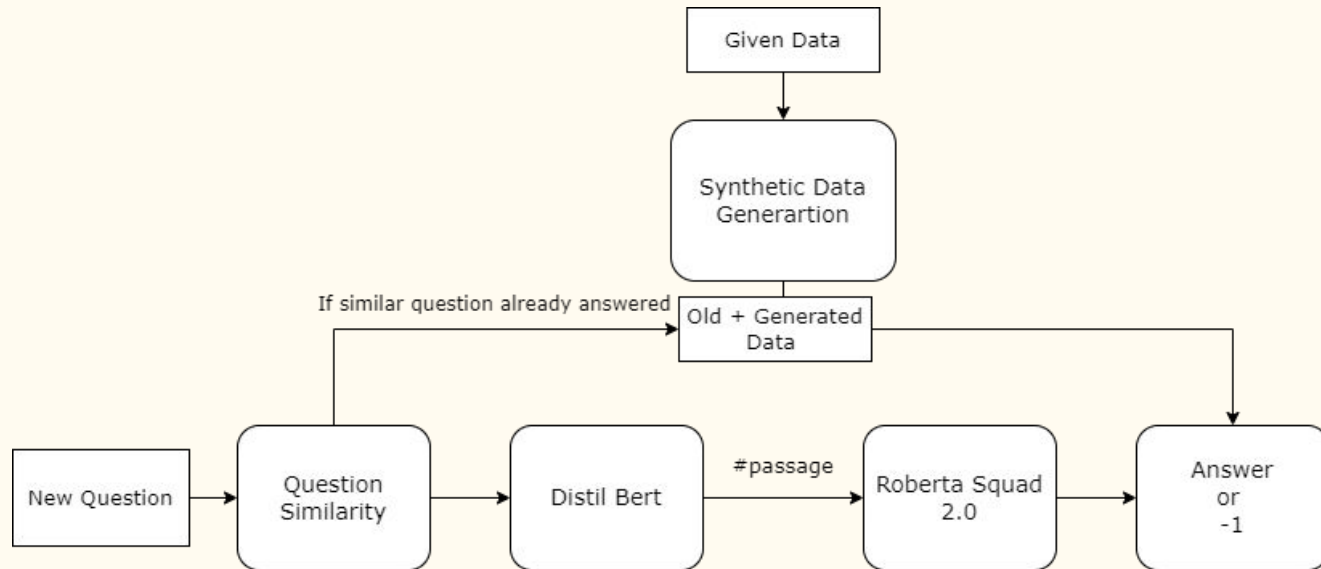
Next task is to point out the exact answer to the question from the passage retrieved...

- RoBERTa has low inf. Time and good performance as well.
- Tiny-BERT has least inf. Time but performance is compromised.
- We choose RoBERTa, as the performance gained is great, with very little compromise in time.



Avg Model Runtime per question: 0.5 seconds

Conclusion: The Final Pipeline



Avg Model Runtime per question: 0.3 seconds