Question/Answering using Encoder and Decoder Transformers

Authors:
Paresh Karan
Kota Mohith
Desai Saharsh

Date: 04/19/2023

Problem Addressed

"What and Why"

Question answering is a challenging task due to the complexity, diversity, and ambiguity in information.

Traditional search engines or information retrieval systems may not be effective for answering complex questions.

Existing question answering systems have limitations in terms of accuracy, efficiency, or domain-specificity.

A specialized question answering system that can provide accurate and efficient answers to complex questions can assist professionals in their decision-making processes.

How does your work relate to existent work?

Previous work in question answering has focused on developing rule-based, knowledge-based, and machine learning-based systems.

Deep learning-based approaches using BERT, BioBERT, and T5 have shown promising results in natural language processing tasks, including question answering.

Our work builds upon existing research by using these state-of-the-art models to develop a more robust and effective question answering system.

Approach

Encoder Models BERT and BioBERT Transformer

- Train Test Split
- Answer End Token Extraction
- Tokenization - Context, Question
- Data Preparation using DataLoader
- Training and Fine-tuning the model
- Save the model for future use
- Extract answers from context using start and end token index of model generated answers
- Evaluate the model using Cosine Similarity on Embeddings
- Using Human-in-the-loop GPT-3.5 (ChatGPT) for optimal results attained by the model

Encoder - Decoder Model T5 Transformer

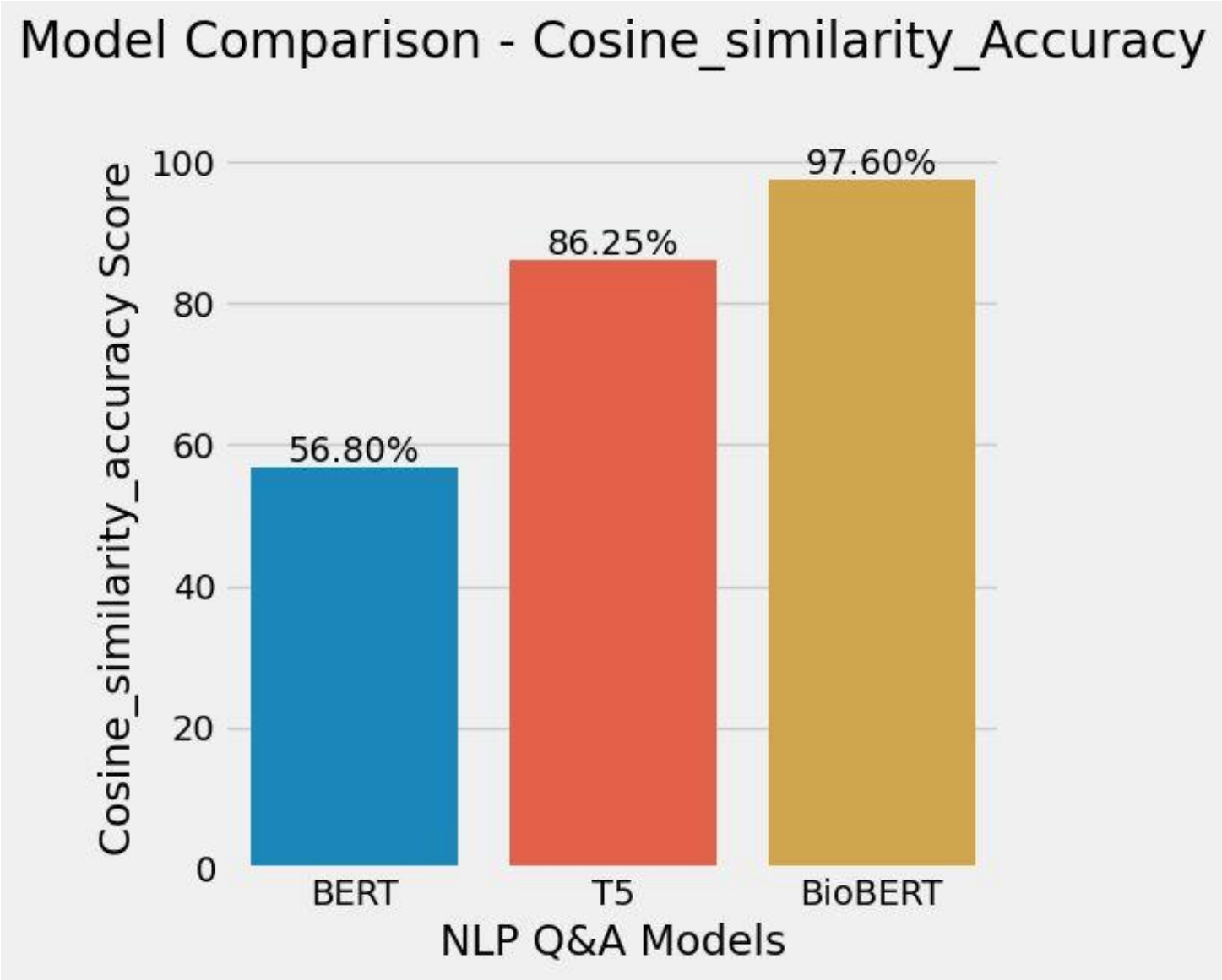
- Train Test Split
- Tokenization - Context, Question
- Data Preparation using DataLoader
- Training and Fine-tuning the model
- Save the model for future use
- Generating answer (word by word) by setting beams to select best generated candidate sequence with highest probability
- Evaluate the model using Cosine Similarity on Embeddings
- Using Human-in-the-loop GPT-3.5 (ChatGPT) for optimal results attained by the model

Results

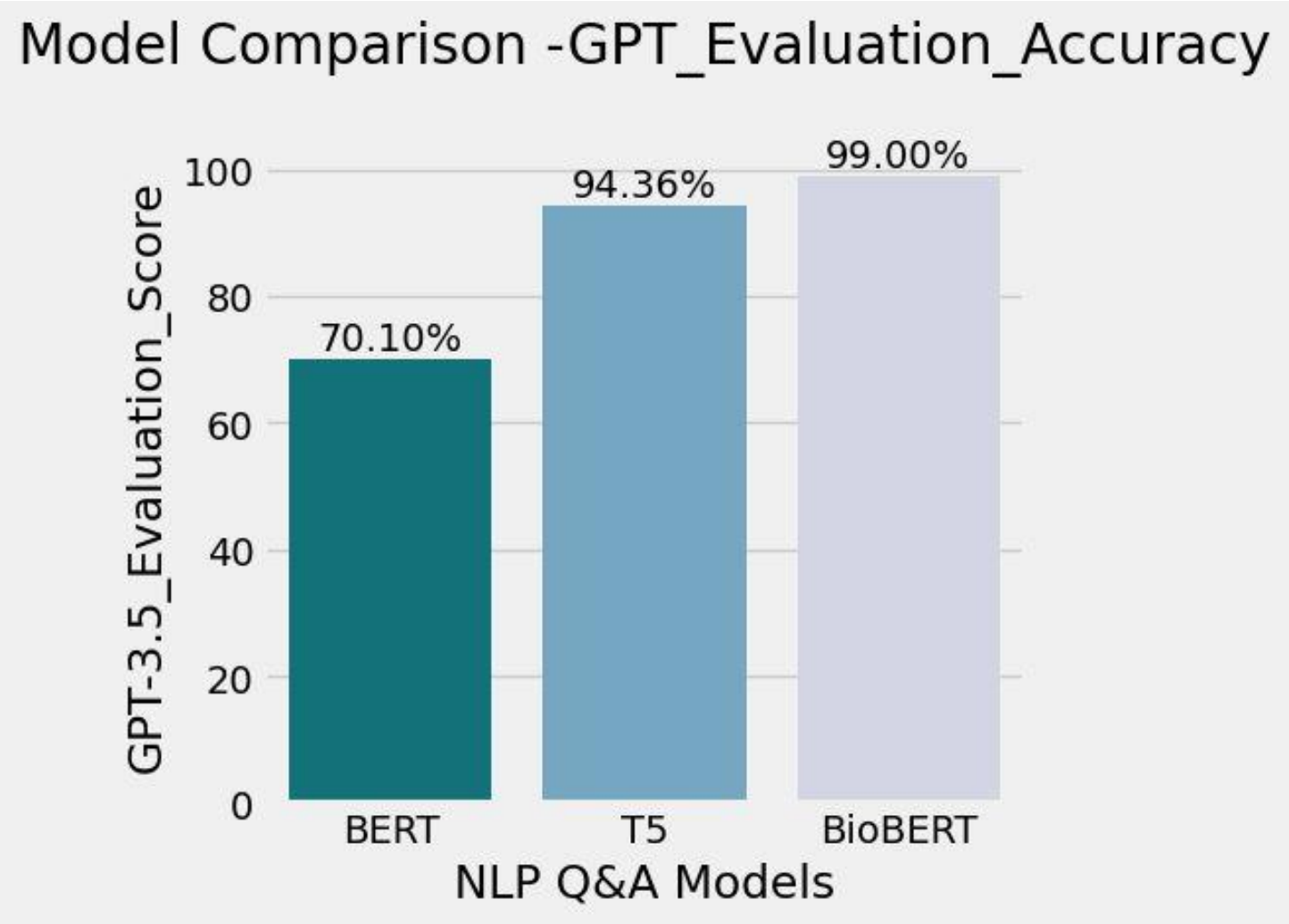
Evaluation

- The dataset we used Stanford Question Answering Dataset (SQuAD) with 87,599 instances is a popular dataset used for training and evaluating question-answering models in natural language processing (NLP).
- Performance measures used for Evaluation - Cosine Similarity, Human-in-the-loop GPT3.5 Accuracy.

Cosine Similarity



GPT3.5 Evaluation



Conclusion

Main takeaway points

- How Encoder, Encoder-Decoder Models are used in Question Answering systems
- How Human-in-the-loop Evaluation helps to check how well the model predicts on the dataset.

- What was achieved
  - We achieved high accuracy and efficiency in question answering, with Cosine Similarity of 97.6% and GPT-3.5 Evaluation Accuracy of 99% using BioBERT.

- Recommended Future Directions
  - The system can be extended to include multi-modal inputs such as images, videos, and audio recordings to provide more comprehensive answers to questions.

THANK YOU