Software Requirements Specification (SRS Document)

Assignment Text Summerization App

By (Ankit Singh)

Table of Contents

1. INTRODUCTION

1.1 PURPOSE

The purpose of this document is to build an online summarization app that can differentiate between two models and gives the best accuracy based on newest technology.

1.2 Approach

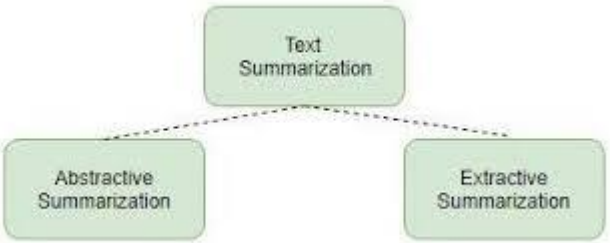
The approach to build this app is independent and experimental though mbd organisation gave freedom to go with innovation that's why we have not used the general approach or previous old models(common) .

2. Description

2.1 Theory about different modeling and approach

There are two approaches to text summarization:

. Extractive approaches
. Abstractive approaches



Some Extractive approaches And Models

- **Gensim**
- **Frequency based method**
- **Sumy librabry**
- **Other library like nlltk,spacy etc.**

Some Abstractive approaches And Models

- **Bart Transformer based model**
- **Seq2seq model**

- **Bert encoder model**
- **Other transformer based architechture.**

2.2 Which Approach and model we selected?

So, we had the opportunity to go with old school general models which mostly organisation uses in or to try some experimental approach and innovate new ideas .

WE TOOK THE DUAL MODEL APPROACH:

1.Fast Text(for extractive method)
2. BART (transformer based model)

1.Fast Text(for extractive method):

This approach is new in the field of nlp to train unsupervised model with a custom dataset. **This is specifically my implementation and idea to create such kind of model,I love challenges and Risks.**

Later combine with fast text word embeddings(which we have trained on our dataset) and find out the cosine similarity with the input article and Each sentence.

Select the best cosine similarity sentence and make a summary.

Advantages of fast text approach over other general models:

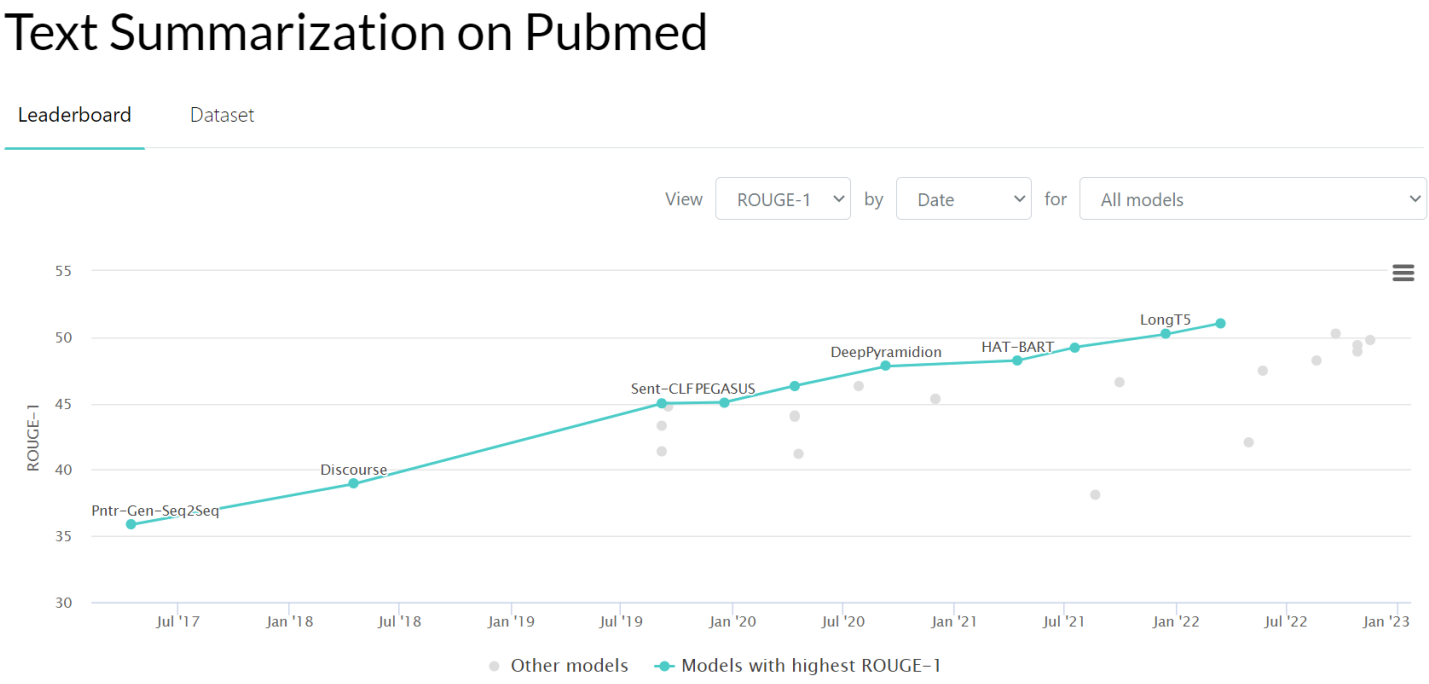
1. Solve the out of vocabulary problem.
2. Solve the misspelling based selection.
3. Word embeddings are more related to domain than general public trained models.
4. Faster To implement than other deep learning projects.

2. BART (transformer based model):

To have also implement the abstractive approach to see the comparison between models.

Advantages of BART (transformer based model) other general models:

1. Highely accurate.
2. Trained on large corpora by facebook.
3. Gives the best crux of the whole article.
4. Can Fine TUNE WITH OUR NEED.

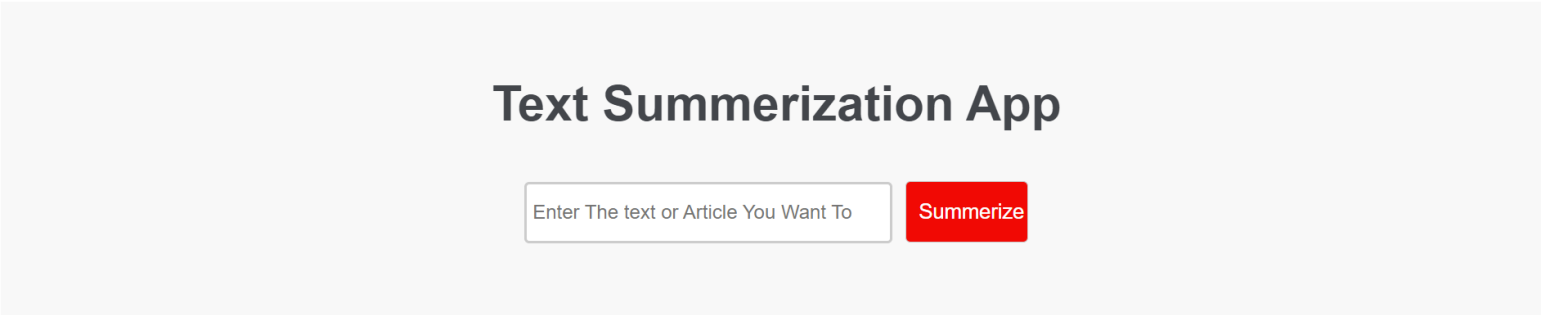


Clearly we can see how popular is bart model specially for medical data.

3. System Features and Requirements

3.1 Dataset

We selected a medical (cancermine) dataset just to find out the accuracy of word embeddings we can generate through fast text.



Input Text : "AI is widely used in the field of healthcare. Companies are attempting to develop technologies that will allow for rapid diagnosis. Artificial Intelligence would be able to operate on patients without the need for human oversight. Surgical procedures based on technology are already being performed. Artificial Intelligence would save a lot of our time. The use of robots would decrease human labour. For example, in industries robots are used which have saved a lot of human effort and time. In the field of education, AI has the potential to be very effective. It can bring innovative ways of teaching students with the help of which students will be able to learn the concepts better. Artificial intelligence is the future of innovative technology as we can use it in many fields. For example, it can be used in the Military sector, Industrial sector, Automobiles, etc. In the coming years, we will be able to see more applications of AI as this technolog is evolving day by day. "

Summary with Extractive Method (Self trained model)

Artificial Intelligence would be able to operate on patients without the need for human oversight.It can bring innovative ways of teaching students with the help of which students will be able to learn the concepts better.Artificial intelligence is the future of innovative technology as we can use it in many fields.In the coming years, we will be able to see more applications of AI as this technology is evolving day by day.In the field of education, AI has the potential to be very effective.

Summary with Abstractive Model(Deep Learning Transformer(BART) based model)

Artificial Intelligence is the future of innovative technology as we can use it in many fields. It can be used in the Military sector, Industrial sector, Automobiles, etc. In the coming years, we will be able to see more applications of AI.

This concludes that our Self Trained model is predicting as good as deep learning BART model!

This Live project is coded and reserved by Ankit singh

Just see How fast text model is selecting and the giving preferences to the medical(technical) sentences shows the embeddings it is generating from our dataset. Here it gave preference to the sentence which has keyword “**Patient”** though the article provided to the model is on artificial intelligence.

3.2 Packages

1. **All dependencies are mentioned in requirement.txt file**
2. **Full software(this app) has a zip file that contain requirements.text,ipynb notebooks,model.py files,dataset file,trained models, all are uploaded to ankit singh github profile.**
3. **Other system requirements are general for transformer based model.**

4.Scope/Improvements/Upgrade

- **Data:** Dataset was too small,if we train on big dataset then word embeddings will have more contextual relations
- **Time:** Main thing was shortage of time,with this model a lot of things can be done,a lot of fine tuning,vectorization ,epochs,training.This model has lot of scope for improvements
- **Hybrid Modeling:** We can also go with combination of models to achieve more accuracy.
- **User Interface:** ofcouse user interface can be changed according to the need with latest front end technologies.Flask and fast api are the main avantage here.