

GROUP 49

Automatic Tweet Summarization

Under the guidance of:
T A Sumesh

Abid Ali KP
B180466CS

OVERVIEW:

1. Introduction
2. Problem Statement
3. Literature
4. Dataset
5. Design
6. Results
7. Conclusion

1. INTRODUCTION

- Natural language processing (NLP) is concerned with giving computers the ability to understand text in much the same way human beings can.
- Automatic Text Summarization is the automated process of generating a concise and meaningful summary from a large text of data.
- Generating a summary with the important phrases in the source text is called Extractive Summarization.
- Generating a summary by understanding the concept in the source text in a more natural language is called Abstractive Summarization.
- Twitter is one of the most widely used social media platforms.

2. PROBLEM STATEMENT

- Given a set of tweets on a particular event, we have to generate a summary for these tweets.

3. LITERATURE

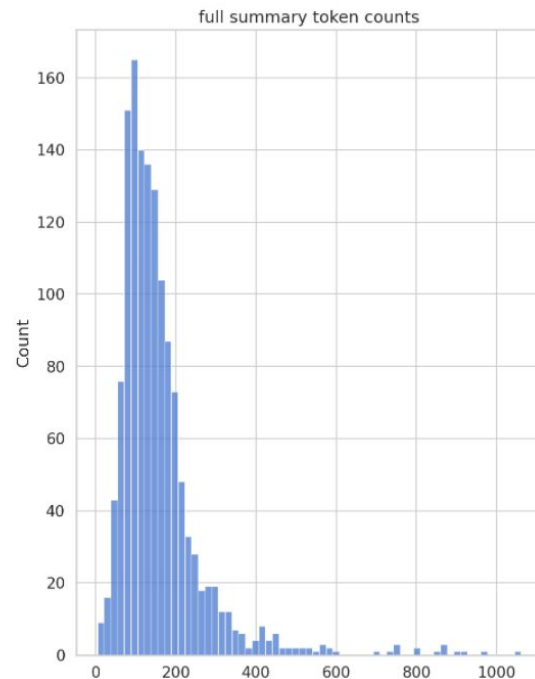
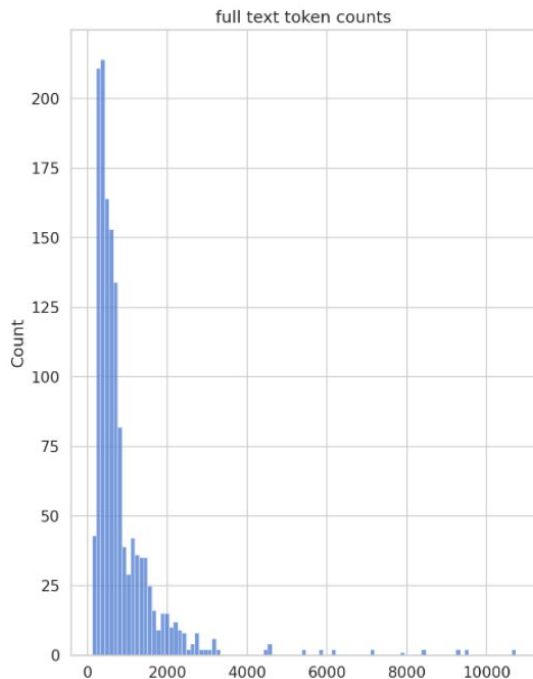
- A recurrent neural network (RNN) is a type of artificial neural network which uses sequential data or time series data.
- Long Short Term Memory (LSTM) networks are a special kind of RNN, capable of learning long-term dependencies.
- The attention mechanism is a part of a neural architecture that enables to dynamically highlight relevant features of the input data, which, in NLP, is typically a sequence of textual elements.
- The Transformer is a novel architecture that aims to solve sequence-to-sequence tasks while handling long-range dependencies with ease.

LITERATURE.....

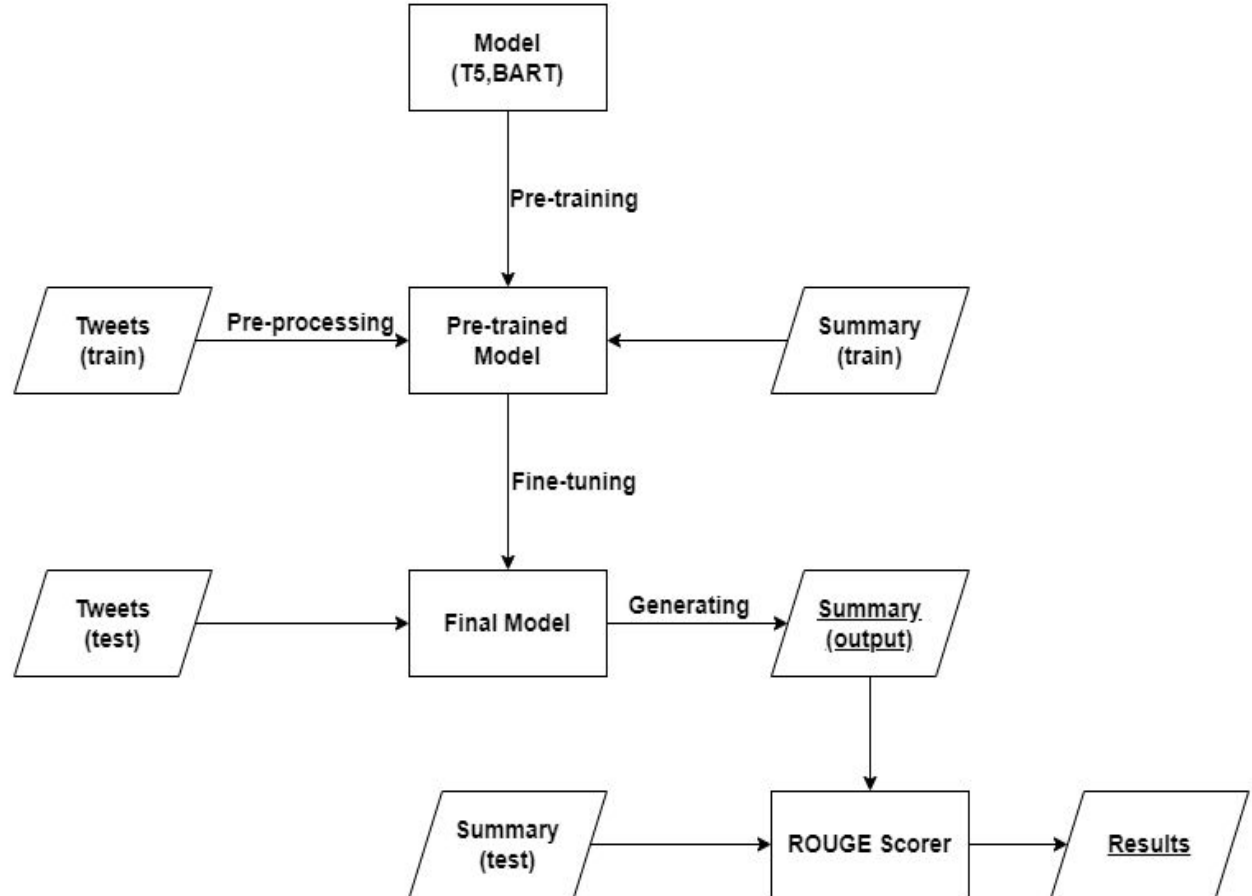
- T5: Text-To-Text Transfer Transformer, is pre-trained on common crawl.
- The model is essentially an Encoder-Decoder Transformer.
- BART is a combination of BERT(Bi-directional encoder) and GPT(left-to-right decoder).

4. DATASET

- TSix - 26 days - Twitter
- Brexit, US election, ISIS, SS Note 7, Nobel prize and SpaceX



5. DESIGN



DESIGN.....

- Tokenization is the process of converting the sentences to a list of tokens.
- Pre-processing: URLs, Mentions, Hashtags, Emojis, Smileys does not convey any meaning to the sentence. Hence they are removed.
- Pre-training: Training a model with one task to help it form parameters that can be used in other tasks. They “learn” the grammatical structures and semantics of a language.
- Fine-tuning: Process of re-training a pre-trained language model using our own custom data. Weights of the original model are updated in this step.
- Generating Summary: The output from the model, a set of tokens, is decoded with the same tokenizer and words are joined to get the summary.

6. RESULT

- ROUGE-N measures unigram, bigram, trigram and higher order n-gram overlap.
- ROUGE-L measures the longest matching sequence of words using LCS.
- Recall score tells us the number of overlapping n-grams between reference summary and model generated summary with respect to total number of n-grams in reference summary.
- Precision score tells us the number of overlapping n-grams between reference summary and model generated summary with respect to total number of n-grams in model generated summary.
- F-measure is calculated by using recall and precision scores.

Everything you need to know about SpaceX's Falcon 9 explosion #Facebook #android #app go to. Everything you need to know about SpaceX's Falcon 9 explosion. Did SpaceX land on Mars yet? We'll tell you when it happens. Subscribe to Elon Musk News. is working on building your home on Mars. According to it will be made of glass domes:. SpaceX CEO Elon Musk dishes about mission to Mars - Upstart :. A fan designed a living quarters concept for Mars spaceship - and we can't wait to climb aboard. If this is what Elon Musk's Mars spaceship will look like on the inside, we're on board via. If this is what Elon Musk's Mars spaceship will look like on the inside, we're on board. SpaceX founder Musk endeavors to find perfect starship name. Episode 3 added to the "Mars Mission 1" films playlist. Everyone's going to Mars with are you?. #ElonMusk unveiled his plan for colonizing #Mars and it is crazy. Can't wait to see the future! #SpaceX. We're going to Mars.... this is what Elon Musk's Mars spaceship will look like on the inside Could an astronaut's corpse bring life to another world? Fascinating post h/t on HN #spacex #mars #seti. SpaceX Founder Elon Musk Explains His Mars Colony Plan to Reddit. Teslas are self-driving. SpaceX will put humans on Mars. This is incredible come-up.. Dragon on its way to Falcon on its way home. We'll start a cadence of sending Dragons to Mars in two years. Will be like a train leaving the station. Intense 360° video takes you for a quick jaunt to Mars via. SpaceX made some awesome travel posters for Mars. Gear up for MARS with this prequel, produced by Variable and colored by. Elon Musk on 'Mars Base Alpha;' plan to test huge fuel tank for the trip - CNET. Florida gives \$5M to SpaceX to help launch crews - USA TODAY. Florida gives \$5M to SpaceX to help launch crews. Florida gives \$5M to SpaceX to help launch crews #TTiNfo. Thanks to for putting their faith in us. Check out Nerdy Show Microsode: Journey to Mars - NASA, SpaceX, and Mars One from The Nerdy Show Network. Leave ship on Mars as habitat. Elon Musk Talks SpaceX Mars Colony Ships and More in Reddit AMA.

Generated Summary:

#SpaceX CEO Elon Musk dishes about mission to Mars - Upstart :... Everything you need to know about SpaceX's Falcon 9 explosion #Facebook #android #app go to.. SpaceX founder Musk endeavors to find perfect starship name.. We're going to Mars.... this is what Elon Musk's Mars spaceship will look like on the inside.... Could an astronaut's corpse bring life to another world?.

Reference Summary:

Elon Musk Talks SpaceX Mars Colony Ships and More in Reddit AMA.. SpaceX CEO Elon Musk dishes about mission to Mars - Upstart :... SpaceX Founder Elon Musk Explains His Mars Colony Plan to Reddit.. If this is what Elon Musk's Mars spaceship will look like on the inside, we're on board via.. Everything you need to know about SpaceX's Falcon 9 explosion #Facebook #android #app go to..

RESULT.....

Table 5.1: ROUGE Metric Scores using Beam Search

Transformer	Metric	<i>Precision</i>	<i>Recall</i>	<i>F-Measure</i>
T5	Rouge-1	0.561	0.700	0.598
	Rouge-2	0.485	0.593	0.514
	Rouge-L	0.518	0.639	0.550
BART	Rouge-1	0.589	0.696	0.618
	Rouge-2	0.516	0.605	0.540
	Rouge-L	0.544	0.639	0.570

RESULT.....

Table 5.2: ROUGE Metric Scores using top-p, top-k Sampling

Transformer	Metric	<i>Precision</i>	<i>Recall</i>	<i>F-Measure</i>
T5	Rouge-1	0.564	0.709	0.604
	Rouge-2	0.488	0.603	0.520
	Rouge-L	0.515	0.639	0.548
BART	Rouge-1	0.601	0.710	0.630
	Rouge-2	0.528	0.620	0.552
	Rouge-L	0.558	0.654	0.584

RESULT.....

Table 5.3: Comparison with different transformers using F-Measure

Metric	<i>T5</i>	<i>BART</i>	<i>GPT2</i>	<i>BERT</i>
Rouge-1	0.604	0.630	0.641	0.701
Rouge-2	0.520	0.552	0.495	0.513
Rouge-L	0.548	0.584	0.506	0.661

7. CONCLUSION

- T5 and BART are two of the best transformer models which achieves state of the art results for various tasks.
- Both were able to generate coherent summaries.
- Our models were able to achieve comparable results using BART with top-p, top-k sampling.



THANK YOU.