

Text summarization problem

DSTI : Deep Learning with Python

Mayank Bhandari
Ouc-Houang Fogoum Philippe Jacques
Ramesh Singh

November 24, 2023

1 Introduction

Text summarization is a crucial aspect of natural language processing, aiming to condense lengthy texts while retaining essential information. In this project, the focus was on utilizing the google/pegasus cnn dailymail model for abstractive summarization.

2 Objective

The primary goal of the project was to automatically generate concise and coherent summaries for given texts using the abstractive summarization capabilities of the selected model.

3 Methodology

3.1 Model Selection

The google/pegasus-cnn dailymail model was chosen for its specialization in abstractive summarization. This model has been pre-trained on a diverse dataset, including the CNN/Daily Mail dataset, making it suitable for a wide range of summarization tasks.

4 Tokenization

we employed a fine-tuned Pegasus model tokenizer to break down input text into meaningful units. This tailored tokenizer, designed for optimal compatibility with the Pegasus model, enhances the text summarization pipeline's effectiveness, ensuring coherent and meaningful abstractive summaries.

4.1 Evaluation Metrics

The model's performance was evaluated using standard ROUGE scores, including "rouge1," "rouge2," "rougeL," and "rougeLsum." These metrics provided a comprehensive assessment of the quality of the generated summaries compared to human-authored reference summaries.

5 Results

5.1 Model Performance Evaluation

For the model evaluation, we are using ROUGE score.

Rouge Names	Result
rouge1	0.015465
rouge2	0.000297
rougeL	0.015503
rougeLsum	0.015514

Dialogue:

Ollie: Hi , are you in Warsaw

Jane: yes, just back! Btw are you free for diner the 19th?

Ollie: nope!

Jane: and the 18th?

Ollie: nope, we have this party and you must be there, remember?

Jane: oh right! i lost my calendar.. thanks for reminding me

Ollie: we have lunch this week?

Jane: with pleasure!

Ollie: friday?

Jane: ok

Jane: what do you mean " we don't have any more whisky!" lol..

Ollie: what!!!

Jane: you just call me and the all thing i heard was that sentence about whisky... what's wrong with you?

Ollie: oh oh... very strange! i have to be carefull may be there is some spy in my mobile! lol

Jane: dont' worry, we'll check on friday.

Ollie: don't forget to bring some sun with you

Jane: I can't wait to be in Morocco..

Ollie: enjoy and see you friday

Jane: sorry Ollie, i'm very busy, i won't have time for lunch tomorrow, but may be at 6pm after my courses?this trip to Morocco was so nice, but time consuming!

Ollie: ok for tea!

Jane: I'm on my way..

Ollie: tea is ready, did you bring the pastries?

Jane: I already ate them all... see you in a minute

Ollie: ok

Reference Summary:

Jane is in Warsaw. Ollie and Jane has a party. Jane lost her calendar.

They will get a lunch this week on Friday. Ollie accidentally called Jane and talked about whisky.

Jane cancels lunch. They'll meet for a tea at 6 pm.

Model Summary:

Jane is back in Warsaw. She lost her calendar. She's going to meet with Ollie on Friday. She'll bring some sun with her.

Figure 1: Text Summary

5.2 Key Findings

The google/pegasus-cnn dailymail model consistently produced coherent and relevant summaries across a diverse set of input texts, as evidenced by the robust ROUGE scores.

6 Conclusion

In conclusion, the integration of the google/pegasus-cnn dailymail model, fine-tuned and evaluated using ROUGE scores, proved successful in addressing the text summarization task. The abstractive

summarization approach enhances the model's ability to distill essential information from input texts, making it a valuable tool for a variety of applications.

7 Project Repository

- [Mayank Bhandari](#)
- [Ouc-Houang Fogoum](#) [Philippe Jacques](#)
- [Ramesh Singh](#)