Individual Project Report

News Article Summarization

Group-02

Meet Daxini

DATS-6312

Natural Language Processing

Dr. Amir Jafari

05/01/2024

Contents

1 Introduction

Making News summaries using transformers

2 Individual Work

1. Doing basic eda of the dataset
2. Scripts to fine tune bart-large-xsum with cnn_dailymail dataset
3. Tried to use llama3 to generate summary
4. Created the app using streamlit with functionalities of external api calling and webscraping with external apps. As the ebscraped data can be of any size and bard had limit of 1024 tokens at a time I also added a function to do summary of 1024 tokens at a time.
5. The percentage of the code that I found from the internet is around 60%

3 Results

To evaluate the effectiveness of my models, I analyzed their performance using rogue scores

Metric	**Score**
ROUGE-1	0.440234
ROUGE-2	0.20731
ROUGE-L	0.303239

Table 1: ROUGE metrics for the CNN-Daily Mail Trained Transformer fine-tuned on bart-large-xsum.

4 Conclusion

This project demonstrates the effectiveness of sequence-to-sequence transformer models for news article summarization. The CNN-Daily Mail model's higher ROUGE scores indicate its superior summarization capabilities.

5 Future Enhancements

1. Using open source llms like LLAMA3 and fine tuning it to be able to use external function that can search and scrape the data from the internet and use rags to have an interactive Q & A with the all the data found from the external sources and give answers with sources listed.

2. Gain a deeper understanding of Encoder & Decoder models with self attention and build one from scratch

6 References

1. Pytorch Documentation.
2. TensorFlow Documentation.
3. Huggingface Documentation.
4. Chat Summarizer.
5. Long Document Summarizer.
6. ChatGPT.
7. Gemini.
8. Claude.