Text Summarization using BART MODEL

INTRODUCTION

Text summarization is a vital task in natural language processing (NLP) that aims to condense large amounts of text into shorter summaries while preserving the key information. In this report, we present our approach to text summarization using the BART (Bidirectional and Auto-Regressive Transformers) model. We will discuss the techniques employed and evaluate the model's performance using ROUGE scores. Additionally, we will explore potential applications of text summarization models like BART.

APPROACH

DATASET

Utilized the CNN/DailyMail dataset, which consists of news articles paired with human-written summaries

BART MODEL

Employed the BART model, specifically the "facebook/bart-large-cnn" pre-trained model

BART TOKENIZER

Used the BART tokenizer to prepare the input data for the model, applying truncation and limiting the maximum length of the input text.

TEXT SUMMARIZATION

developed a function to generate summaries using the BART model

MODEL PERFORMANCE

To evaluate the performance of our BART model for text summarization, the ROUGE (Recall-Oriented Understudy for Gisting Evaluation) metric is used.

ROGUE 1 - UNIGRAMS

0.57

ROGUE 2 - BIGRAMS

0.40

ROGUE 3 - LCS

0.51

Input

If you want to create a machine learning model but say you don’t have a computer that can take the workload, Google Colab is the platform for you. Even if you have a GPU or a good computer creating a local environment with Anaconda and installing packages and resolving installation issues are a hassle. Colab is a free Jupyter Notebook environment provided by Google where you can use free GPUs and TPUs which can solve all these issues. Input Text: If you want to create a machine learning model but say you don’t have a computer that can take the workload, Google Colab is the platform for you. Even if you have a GPU or a good computer creating a local environment with Anaconda and installing packages and resolving installation issues are a hassle. Colab is a free Jupyter Notebook environment provided by Google where you can use free GPUs and TPUs which can solve all these issues.

Output

Bart Generated Summary: If you want to create a machine learning model but say you don't have a computer that can take the workload, Google Colab is the platform for you. Even if you have a GPU or a good computer creating a local environment with

APPLICATIONS

News Aggregation:
Summarizing news articles can help users quickly grasp the main points

Document Summarization:
For large documents, automatic summarization can provide executive summaries saving time for readers.

Information Extraction:
Text summarization can aid in extracting key information from lengthy documents, such as legal texts or scientific papers.

Chatbots and Virtual Assistants:
Incorporating text summarization models can enhance chatbots and virtual assistants by generating concise responses based on user queries.

Thank You