Abstractive Summarization Evaluation Report

1. Overview

This study uses a dataset of news items to assess two pre-trained summarisation models, T5-small and BART-large-CNN. Finding out how well these models perform in comparison to summaries provided by humans is the aim.

2. Models Employed

T5-small: A portable model with text summarisation capabilities.
A larger model designed especially for news story summarisation is the BART-large-CNN.

3. The dataset

Articles and their human-written summaries are included in the dataset, which is a JSON file. For evaluation, we used the dataset's initial ten samples.

4. Getting ready

To conform to the models' input restrictions, articles were reduced to 512 tokens. This maintains crucial content while guaranteeing compatibility.

5. Creation of Summaries

Both the T5 and BART models were used to summarise each article. The original human summaries and the produced summaries were contrasted.

6. Assessment

To assess the quality of the generated summaries, ROUGE scores (ROUGE-1, ROUGE-2, and ROUGE-L) were employed. The degree to which the model outputs resemble summaries provided by humans is gauged by these scores.

7. Findings

A table displaying each model's performance on each article contained the scores. This makes it simple to compare the T5 and BART models' performances.

8. Graphics

The performance of the T5 and BART models across the evaluation dataset is shown in the following charts.

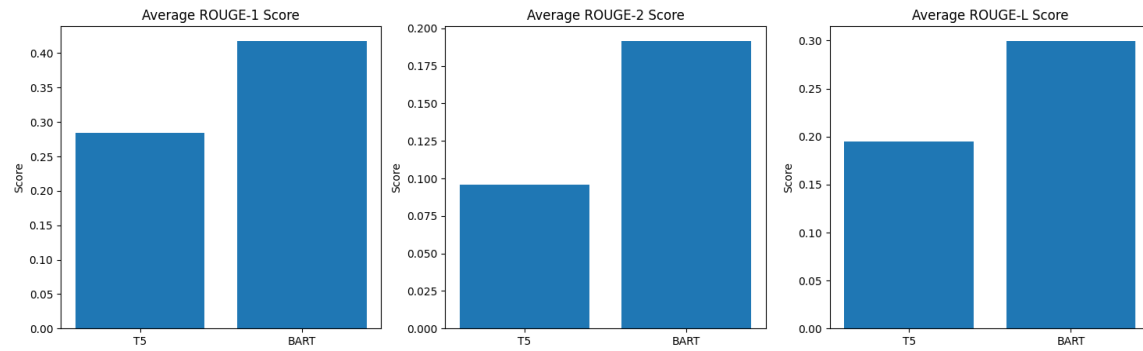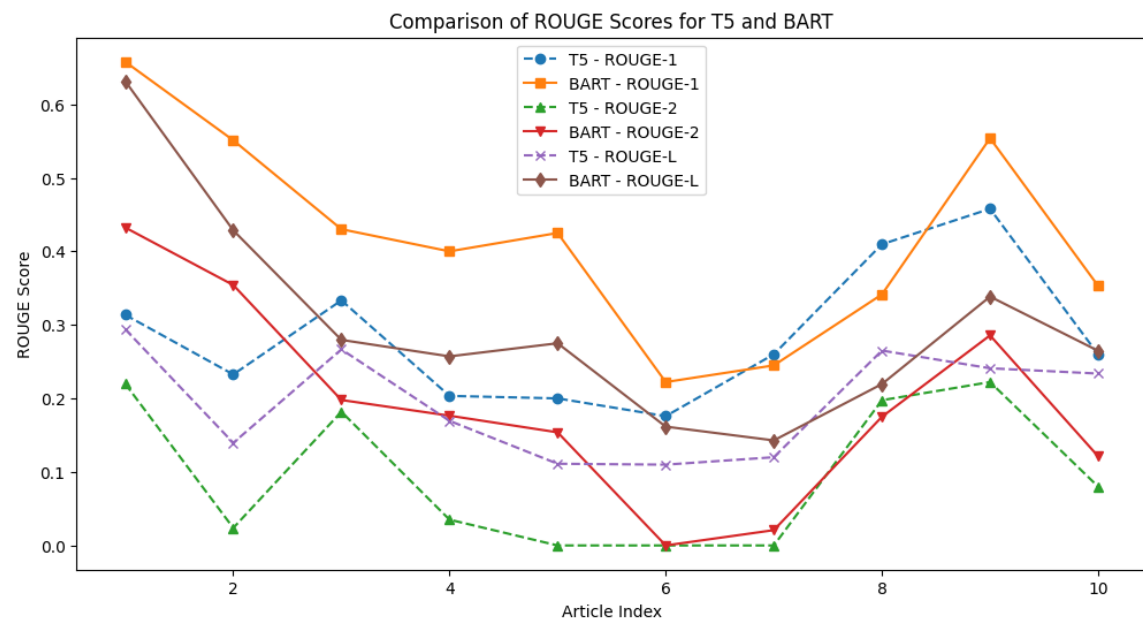Figure 1: Average ROUGE Scores for T5 and BART

Figure 3: ROUGE Scores for Each Article (Detailed Line Chart)

9. In conclusion

According to the study, BART-large-CNN performs better than T5-small on all ROUGE criteria. In terms of both substance and organisation, BART shows greater overlap with summaries that are authored by humans. T5-small offers substantially inferior summarisation quality despite being faster and lighter computationally. The BART model is the recommended choice for applications that want higher-quality summaries. In environments with limited resources, T5 might be helpful.