

Optimizing Scientific Paper Summarization with Fine-Tuned T5 on the ArXiv Dataset

Muhammad Ibtisam Afzal
FA22-BCS-073
05 Oct 2024

Abstract

In today's fast-paced scientific world, the rapid increase in research publications has created a need for better paper summarization tools. The motivation for this study comes from the need to help researchers quickly understand large amounts of scientific information. A major challenge is developing summarization models that stay clear and accurate, even with the complexity of the original material. Earlier methods, including transformer-based models like T5, have shown promise but often struggle with specific subject details and handling large amounts of data.

In this work, we fine-tune the T5-small model on the ccdv/arxiv-summarization dataset to improve the summarization of scientific papers. Our contributions include enhancing the model's ability to identify important ideas and structure within scientific texts, improving summary quality while keeping it efficient. We also use the ROUGE score to carefully evaluate how well the model performs and highlight areas for further improvement.

This research is important because it can significantly reduce the time researchers spend reading papers, making it easier to share and spread knowledge. By improving summarization abilities, our model can help create scalable, automated tools that make the research process faster and more efficient in different fields.

Keywords

Scientific paper summarization, Fine-tuned T5 model, Automated summarization, ArXiv Summarization dataset, ROUGE score evaluation, Research efficiency, Scientific Paper Summarization, Abstractive Summarization.