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 ~~environment~~world, the rapid ~~growth of~~increase in research publications has created a ~~demand~~need for ~~efficient~~better paper summarization tools. The motivation for this study ~~stems~~comes from the need to help researchers quickly ~~comprehend~~understand large ~~volumes~~amounts of scientific ~~literature~~information. A ~~key~~major challenge ~~lies in~~is developing summarization models that ~~maintain both coherence~~stay clear and ~~accuracy, despite~~ accurate, even with the complexity of the ~~source~~original material. ~~Previous approaches~~Earlier methods, including ~~various~~ transformer-based models like T5, have ~~demonstrated potential,~~shown promise but ~~they~~ often struggle with ~~domain~~ specific ~~nuances~~subject details and ~~scalability~~handling large amounts of data.

In this work, we fine-tune the T5-small model on the ~~ArXiv~~ccdv/arxiv-summarization dataset to ~~optimize scientific paper~~ improve the summarization~~.~~ of scientific papers. Our contributions include ~~improving~~enhancing the model's ability to ~~recognize critical concepts~~identify important ideas and structure within scientific texts, ~~enhancing the summarization~~ improving summary quality while ~~maintaining computational efficiency.~~keeping it efficient. We also ~~employ~~use the ROUGE score ~~for a rigorous evaluation of~~ to carefully evaluate how well the model ~~performance, highlighting~~performs and highlight areas for further ~~refinement~~improvement.

~~The significance of this~~This research ~~lies in its potential to substantially~~is important because it can significantly reduce the time researchers spend ~~reviewing literature, fostering more efficient~~ reading papers, making it easier to share and spread knowledge~~ dissemination.~~. By ~~advancing~~improving summarization ~~capabilities~~abilities, our model can ~~contribute to the development of~~help create scalable, automated tools that ~~streamline~~make the research process ~~across various domains~~faster and more efficient in different fields.

Keywords

Formatted: Justified

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