Literature Review

- Akshay Chavan, Gaurav Tejwani, Ananya Asthana

[1]

A. D. Bhargavi and Ch. D. V. S. Rao, "Video Transcripts Summarization using OpenAI Whisper and GPT Model," *Int. J. for Research in Applied Science & Engineering Technology*, vol. 12, no. 3, pp. 1371–1377, Mar. 2024. Link: https://www.researchgate.net/publication/378249741.

Paper Summary

In their paper, Bhargavi and Rao present a pipeline that combines OpenAI's Whisper for transcription and a GPT-based model for summarization. The authors transcribe educational video content using Whisper and then pass the resulting transcripts to a generative LLM for abstractive summarization. The study evaluates the quality of the summaries using 'ROUGE' scores and shows that the Whisper + GPT combination outperforms previous summarization pipelines in terms of accuracy and relevance. The paper highlights challenges with long transcript processing and suggests using a chunking strategy to manage lengthy content. The authors note that Whisper's larger models deliver better transcription accuracy, but complex technical terms and noisy audio remain challenging. They propose refining the summarization model by improving the handling of complex domain-specific terms.

Relevance to our Project

This paper closely aligns with our project's overall pipeline. Like Bhargavi and Rao's approach, our project also involves transcribing lectures and summarizing them using an LLM. Another similarity is that our project plans to use Whisper, like this paper's approach. The insights into using such chunking strategies to handle long transcripts directly apply to our project's summarization component. Also, the evaluation of summary quality using 'ROUGE' scores provides a benchmark that we can adopt for our own use case. This paper's recommendation to refine the handling of domain-specific terms is relevant to us since lecture content is often rich in technical vocabulary, and this can be used as the key domain.

Comparison with our Project Approach

The summarization approach is otherwise similar, but our project's use of FLAN-T5-Large instead of a GPT-based model could result in slightly different summarization quality (depending on the underlying model's training and parameter tuning). Aditionally, our project's plan to include a separate Q&A component introduces an extra layer of complexity, which Bhargavi and Rao's system does not address. However, their work validates the advantages of combining ASR and LLM-based summarization, which our project does not address exactly.

[2]

Z. Kolagar and A. Zarcone, "HumSum: A Personalized Lecture Summarization Tool for Humanities Students Using LLMs," in *Proc. 1st Workshop on Personalization of Generative AI Systems (PERSONALIZE@ACL)*, 2024, pp. 36–70. Link: https://arxiv.org/abs/2401.12345.

Paper Summary

In their paper, Kolagar and Zarcone develop a tool specifically for summarizing lecture transcripts using large language models (LLMs), emphasizing customization and scalability. The system, named HumSum, processes lecture transcripts and generates summaries that can be customized to the user's preferences, such as focusing on specific topics or adjusting the level of formal tone. The authors employ a map-reduce approach using LangChain to split long transcripts into manageable chunks, ensuring that even lengthy content can be processed efficiently. GPT-4 is used to generate summaries, and the system allows taking user feedback to improve future summaries. The paper reports that personalization improves user satisfaction and understanding.

Relevance to our Project

This paper is highly relevant to the summarization component of our project. Its goal of summarizing lecture transcripts using an LLM aligns closely with HumSum's objective. The map-reduce strategy to handle long transcripts is definitely a valuable insight that we can utilize since lectures can be long and complex. Personalization of summaries through user feedback is a unique feature that the project could explore as a future enhancement. We have identified that the personalization element in HumSum could be adapted to our project's Q&A system by customizing answers to the user's query style or academic background, and we can take inspiration from it for our own use case.

Comparison with our Project Approach

The key similarity is the use of an LLM for summarization and the need to manage long transcripts. However, our project focuses on educational content beyond just the field of humanities, while HumSum is specialized for humanities lectures. Unlike Kolagar and Zarcone's model, our project's pipeline includes a QA component that allows users to ask questions based on the summarized content and transcript generated. While HumSum uses GPT-4, our project plans to use FLAN-T5-Large, which may result in differences in language generation style, detail amongst other things.