Enhancing Educational Accessibility through Speech and Language Processing

Xin Cao (120040062)

Abstract

This report proposes the development of an application that uses automatic speech recognition (ASR), natural language processing (NLP), and machine learning (ML) techniques to transcribe spoken lectures and align them with corresponding PowerPoint (PPT) slides and highlights. By segmenting lecture transcriptions according to the PPT outline and highlighting them based on spoken cues and domain knowledge, this tool aims to enhance the educational experience for students: particularly those who rely on written content for learning, wish for quick-learning on a course, or desire to review a course.

1. Introduction

The increasing use of digital technology in education offers unique opportunities for leveraging advanced computational techniques to support learning. This proposal addresses the need for improved accessibility and engagement in educational settings through the application of speech and language processing (SLP) technologies. Students often struggle to correlate live or recorded lectures with their corresponding PPT slides due to lecturing pronunciation, difficulty with foreign language proficiency, lack of pre-knowledge, or simply because they don't have enough time to watch the whole recording. Additionally, identifying key information during lectures that are not text-based presents a challenge, particularly for those with auditory processing issues or those who are not native speakers.

2. Related Work

I found three papers in the field of educational technology exploring the integration of Automatic Speech Recognition (ASR) and highlighting, which are foundational to the proposed system.

2.1 AI-inspired Multilanguage Framework for Note-taking and Qualitative Content-based Analysis of Lectures

An investigation into artificial intelligence-driven multilanguage frameworks highlights the potential of ASR systems to function across different linguistic contexts. This paper supports our work by discussing the integration of ASR in environments with multiple languages, expanding the accessibility of educational content beyond monolingual boundaries.

2.2 A Comprehensive Survey on Process-Oriented Automatic Text Summarization with Exploration of LLM-Based Methods

A foundational paper on process-oriented tasks provides insights into how ASR can be used to structure verbal information into more accessible and manageable forms. This research is relevant to our project as it outlines methods for capturing spoken content and processing it into summarized, organized formats, an essential component of our lecture transcription and highlighting system

2.3 Towards better information highlighting on technical Q&A platform

Ahmed and Shahla (2023) used CNN and BERT architectures to develop automatic information highlighting models and achieved an F1 score of 0.71 and 0.82. This approach may be highly related to when developing our own model to highlight information on our lecture transcript.

These studies collectively underscore the viability and necessity of employing ASR and intelligent text-processing technologies in educational settings. By leveraging these technologies, our system aims to not only transcribe but also segmentize and semantically enrich lecture content, making it more accessible and beneficial for diverse learner populations.

3. Target Population

The primary users of this application will be students and lecturers in higher education who attend lectures that utilize a significant amount of spoken communication and visual aids (PPTs). The application will also benefit educationists or people who are interested in certain topic by granting them the capability of going through official lectures quickly and efficiently.

4. Related Applications

Lecture capture systems: records audiovisual lectures, but don't offer sophisticated integration of speech with visual content.
Enhanced e-book readers: synchronizes audiobook content & text, but limited to pre-recorded standardized content.
ASR-based note-taking tools: provides transcription, but lack contextual awareness and multimodal integration.

5. Proposed Solution

5.1 Technology Overview

- Automatic Speech Recognition (ASR): to transcribe spoken words into text.
- Text Segmentation and Alignment: dividing the lecture transcript according to the PPT outline and correlate spoken words with specific PPT slides based on PPT page information (as an input).
- Multimodal Highlighting Model: uses machine learning to integrate verbal intonations, information-highlighting and automatic text-summarization techniques, field-specific terminology annotating, and visual slide highlights to identify and mark key information in the transcript.

5.2 Multimodal Highlighting Model Design

- Machine learning to detect verbal intonations: train a model that takes the audio of a paragraph of spoken lecture or speech as input, and a few words that are of high intonation or heavily cued by the speaker as output. The model will be capable of classifying the words into different intonation levels. A simple RNN architecture would do, and Word Error Rate (WER) can be used to evaluate the performance of this model.
- Information-highlighting techniques Automatic text-summarization: relating to Ahmed and Shahla's work we may use their BERT model and finetune them on some of our own education-related datasets. We may then combine this step with the keywords appearing a summarized edition of the transcript, hence obtain more important keywords.
- Field-specific terminology annotating: we obtain the topic or field of the lecture and invoke GPT's API to list the top keywords or important terminologies in the transcript.
- Visual slide highlights: implement an algorithm that looks for highlighted or bolded or colored words or phrases in the PPT slides. We finally combine all steps above and train them together to find the weights of the four steps.

6. Anticipated Results

6.1 Enhanced Lecture Accessibility

By providing a text-based version of spoken lectures aligned with visual aids, the application will make learning more accessible to students with hearing impairments, auditory processing difficulties, and non-native speakers.

6.2 Improved Information Retention

The ability to highlight key concepts based on a multimodal analysis (main dish of this project) will help students focus on the most important information, potentially improving study efficiency and knowledge retention.

6.3 Increased Flexibility in Learning

With transcripts that accurately reflect the structure and key points of lectures, with a bonus summarization feature, students can review efficiently in a manner that best suits their learning style, whether visual, auditory, or text-based.

7. Conclusion

This report proposes a novel application of speech and language processing (SLP) technologies to solve a significant problem in educational settings. By integrating ASR, NLP, and machine learning, the application not only enhances accessibility and learning effectiveness & efficiency, but also sets a foundation for future advancements in educational technology. Through such innovations, we can better cater to diverse learning needs and maximize learning recourses educational outcomes.

References

Hanlei Jin, Yang Zhang, Dan Meng, Jun Wang, Jinghua Tan (2024). *A Comprehensive Survey on Process-Oriented Automatic Text Summarization with Exploration of LLM-Based Methods*, arXiv:2403.02901v1 [cs.AI]

Munish Saini, Vaibhav Arora, Madanjit Singh, Jaswinder Singh, Sulaimon Oyeniyi Adebayo (2022). *Artifcial intelligence inspired multilanguage framework for note-taking and qualitative content-based analysis of lectures*, Education and Information Technologies, 28:1141-1163 https://doi.org/10.1007/s10639-022-11229-8

Ahmed, Shahla (2023). *Towards better information highlighting on technical Q&A platform*, http://hdl.handle.net/1993/37572

Jurafsky, D., & Martin, J. H. (2022). *Speech and language processing* (4th ed.). Draft available online. Retrieved from https://web.stanford.edu/~jurafsky/slp3/