CSCE 5290: Natural Language Processing

Project Proposal

Title: Integrating Speech Recognition and Text Summarization for Conversational AI.

GitHub link: https://github.com/SaiKiranReddyK28/NLP-Project-Group-9.

Team Members:

Aashish Vinay Vasala (11653925)

Shiny Shamma Kota (11698716)

Sai Kiran Reddy Kancharla(11605339)

Motivation:

In the growing world where AI plays a crucial role in our daily lives like the usage of Alexa, Siri and other Speech recognition applications, It is crucial to enhance the user experience. In order to do so we need proper speech recognition technology. The core problem to solve these, is to create AI conversation systems that can seamlessly recognize and summarize the speech of people, making the communication between machines and humans a natural act. In order to optimize the performance of the machines it is imminent to have the system summarize the user given text into keywords to deliver faster performance of the model, to fit with the short user attention span. This project is driven by an increasing need for conversational AI options that require fast model outputs in different areas such as customer service, voice assistants, and social bots.

By Integrating speech recognition and text summarization, the AI becomes more accessible to the users with disabilities, also facilitates seamless interaction between individuals from different regions to communicate effortlessly. It also helps people multitask, perform efficiently and get more work done in a short duration.

Significance:

The significance of this project is that the model built will allow the conversational AI system to learn speech recognition and text summarization techniques, in order to address the motivation behind the proposal. The model proposes to have speech recognized and have it summarized with optimal parameters. S. Furui and et al [3] have done speech recognition and summarization for spontaneous speeches for instances such as live video presentations, lectures, Team meetings in offices etc. Considering the importance and the potential for a lot of domains, Implementing this vision is a vital part of the development and exploitation of NLP field capability as well as of a successful conversational AI technology.

Objectives:

- Design a meaningful speech recognition model, which translates human speech into text with a high degree of accuracy.

- Apply the text summarization that highlights the main topics and major themes of speech that are transcribed.

- Integrate in the unified solution speech recognition and text summarization algorithms.

- Fine-tune the integrated model to ensure optimal performance and coherence in summarizing the content which is spoken.

- Evaluate the effectiveness of the integrated system through qualitative and quantitative metrics, including accuracy, coherence, and user satisfaction.

Features:

- Technical Characteristics: Utilizing state-of-the-art models such as BERT for speech recognition and text summarization tasks.

- Deliverables: A fully functional conversational AI system capable of understanding and summarizing spoken language.

- Milestones: Development of individual components (speech recognition, text summarization), integration of components, fine-tuning of the integrated model, and evaluation of system performance.

Dataset:

We will utilize two datasets for training and testing our conversational AI system:

Reddit Dataset: This dataset consists of conversational threads from various topics extracted from Reddit. It provides a rich source of conversational data for training both the speech recognition and text summarization components of our system. Preprocessing steps will include data cleaning, tokenization, and alignment of speech transcripts with corresponding summaries.

Amazon Product Reviews Dataset: This dataset contains customer reviews for various products available on Amazon. It will be used to evaluate the effectiveness of our text summarization module specifically in summarizing product reviews into key insights. Preprocessing steps will involve filtering irrelevant data, extracting relevant reviews, and preprocessing the text for summarization.
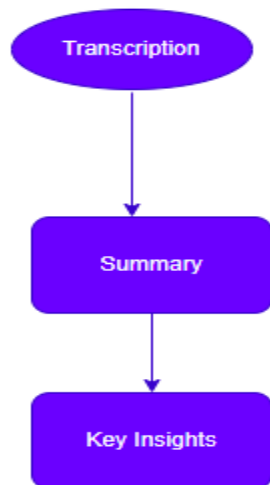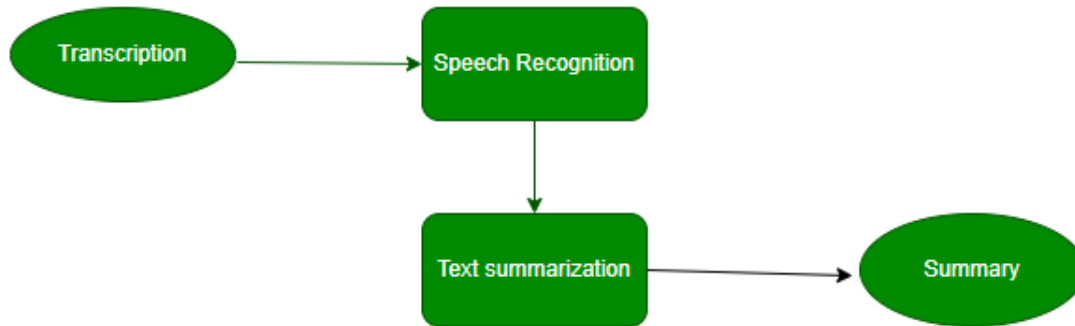
Visualization:

Workflow Diagram:

Speech Recognition Module Workflow:
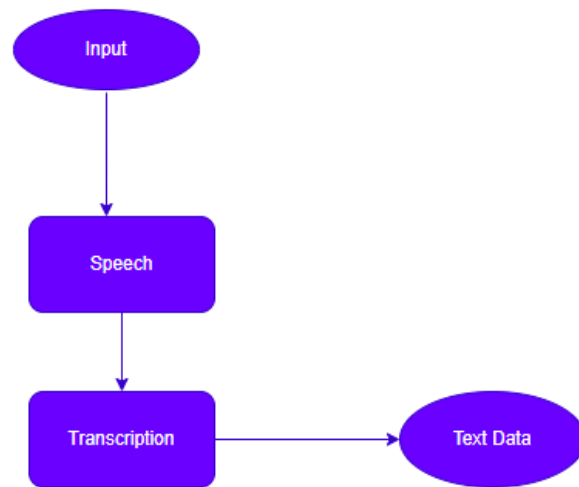
Text Summarization Module:

Integration of Modules:

Tables:

Evaluation Metrics Table:

Metric	Description
Accuracy	Percentage of correctly transcribed speech
Coherence	Degree of logical consistency in summaries
User Satisfaction	Subjective assessment of user experience

Dataset Statistics Table:

Dataset	Sources	Preprocessing Steps
Reddit Dataset	Reddit	Data cleaning, tokenization, alignment
Amazon Product Reviews	Amazon	Filtering, extraction, text preprocessing

Team Members and Work Divison:

Aashish Vinay Vasala: Data Cleaning, Filtering

Shiny Shamma Kota: Tokenization, Extraction

Sai Kiran Reddy Kancharla: Alignment, Text Preprocessing.

References:

1. Dataset 1: https://www.kaggle.com/datasets/thedevastator/the-meta-corpus-of-datasets-the-reddit-dataset.

2. Dataset 2: https://www.kaggle.com/datasets/saurav9786/amazon-product-reviews

3. S. Furui, T. Kikuchi, Y. Shinnaka and C. Hori, "Speech-to-text and speech-to-speech summarization of spontaneous speech," in IEEE Transactions on Speech and Audio Processing, vol. 12, no. 4, pp. 401-408, July 2004, doi: 10.1109/TSA.2004.828699.