Voice Assistant for Indic languages

Sai Kasyap Kamaraju

HT23RESCH04001

Abstract

Building a voice assistant for low resource Indian languages like Hindi/Sanskrit is challenging task. In this project, it is proposed to develop a voice assistant (ignoring the hardware requirements) for Hindi/Sanskrit using a small Heritage Text corpus and leveraging pre-trained models for automatic speech recognition (ASR), language modeling (LM), and text-to-speech (TTS). The first step in developing the voice assistant is to collect a small corpus of Hindi/Sanskrit ancient text. This corpus will be used to Finetune and evaluate the ASR and LLM models. The pre-trained/ Fine-tuned ASR models like IndicWav2vec, Whisper will be used to recognize the speech query of the user. Next, a Large Language Model is used so as to capture the linguistic and contextual information of Hindi/Sanskrit text. Pre-trained model like Flan-t5-xxl is used and fine-tuned on the collected corpus to generate text that sounds natural and coherent in Hindi/Sanskrit. Once the ASR and LM models are in place, those are integrated into a dialogue management system so as to understand user queries and generate appropriate responses. Open-source libraries like langchain, Rasa are used develop the dialogue management system and train it on the collected corpus. Finally, it is integrated with a pre-trained TTS model like FastPitch and HiFiGAN (API) to generate natural-sounding speech output for the user queries. The developed voice assistant can be evaluated using a combination of objective metrics like word error rate (WER) and subjective evaluations like user feedback to ensure that it provides accurate and natural responses to user queries. In summary, this project aims to develop a voice assistant for Hindi/Sanskrit using a small corpus and leveraging pre-trained models for ASR, LLM, and TTS. The voice assistant will be evaluated using a combination of objective and subjective metrics to ensure high accuracy and naturalness.