Audio Pipeline for Sandalwood Cultivation Knowledge Retrieval System

Comprehensive Solution Report

Problem Description

- D evelop an ASR mod el for Kannad a language audio related to sandalwood cultivation.
- Hand le colloquial language and noise in audio data.
- Build a question-answer ing system for Hin speaking users to query the Kannada audio corpus.
- Pr ovid e r esponses in Hind i using tr ansla and TTS.

Project Objectives

- 1. Fine-tune an ASR model for colloquial Kannada speech.
- 2. Create a searchable repository of transcripts for sandalwood cultivation audio.
- 3. Implement a QA system that handles Hindi queries and retrieves answers from Kannada audio data.
- 4. Build a user-friendly interface for seamless querying and response retrieval.

Solution Architecture

1. User pr ovid es a Hind i voice input.
2. ASR conver ts the Hind i voice input to text.
3. Hind i text is tr anslated to Kannad a using MarianMT.
4. Sear ch engine r etr ieves r elevant Kanna transcript segment.
5. Kannad a answer is tr anslated back to Hi and converted to audio using TTS.
6. R esponse is played back to the user in Hindi

Technologies Used

- Backend : Flask (Python)
- iOS App D evelopment: Xcod e, Swift, SwiftUI
- ASR M od el: Whisper , Wav2Vec 2.0
- Tr anslation M od el: M ar ianM T (Hind i-Kannada)
- Sear ch Engine: Elasticsear ch, FAISS
- TTS: AVSpeechSynthesizer (Swift)
- Hosting: AWS EC2, Fir ebase

Dataset Details

- Aud io d ata scr aped fr om YouTube, r ela sandalwood cultivation in Kannada.
- Includ es public r ecor d ings with minor no and colloquial language.
- Pr epr ocessing involved noise r ed uction segmentation using librosa.

ASR Model Development

1. Data Preparation: Noise reduction and normalization using librosa.
2. Model Selection: Whisper and Wav2Vec 2 chosen for robust performance.
3. Fine-tuning: Customized on Kannada dataset for colloquial speech.
4. Evaluation: Achieved Word Error Rate (WER) of 12.5%.

Question-Answering System

1. Hind i ASR captur es user quer y.
2. Quer y is tr anslated to Kannad a using MarianMT.
3. Elasticsear ch r etr ieves r elevant tr ansc segments.
4. Kannad a answer is tr anslated to Hind i a read aloud using TTS.
5. Achieved quer y r etr ieval accur acy of 8

User Interface Design

- Simple iOS app with a micr ophone button for Hindi voice input.
- D isplays r ecognized question and r espon text.
- Pr ovid es Hind i aud io output for user convenience.

Testing and Evaluation

- ASR Model Accuracy (WER): 12.5%
- Translation Accuracy: 90%
- Query Retrieval Accuracy: 85%
- User Satisfaction (Survey): 92%

Conclusion and Future Work

- Successfully captur ed ind igenous knowled and provided an accessible query system.
- Futur e Enhancements:
- - Impr ove ASR accur acy with a lar ger d
- - Extend suppor t for other r egional langu (Tamil, Telugu).
- - Scale backend ser vices for higher user concurrency.

References

1. Whisper Model Documentation: OpenAI Whisper
2. MarianMT Model: Hugging Face Marian
3. Elasticsearch Documentation: Elasticsearch Guide