Regional Voice Assistant

P Deepak Reddy, Chirag Rudresh, Adithya A S, Dr. Uma D
Department of Computer Science and Engineering, PES University

Problem Statement

Developing a voice assistant that comprehends multilingual voice navigation queries by recognizing the various languages used in a multilingual sentence.

Background

MFCC is the state of the art feature extraction method used for speech processing.
The performance of i-vector based approaches will start to saturate when the training data starts to exceed certain limit hours.
Variants of the LAS model which can be used to better recognize monolingual audio sentences without the explicit mentioning of the language being used.

Dataset and features/project requirements/Product Features

Dataset contains all the possible multilingual translations (POS tagged and recorded) for each monolingual sentence.
Both Audio and textual dataset were required for training. The audio data should be in wav format.
The query should contain word utterances with sufficient gaps.

Design Approaches/Methods

Extracted MFCC features from pre-processed audio recording were used for training the Deep Learning model.The alternative faster approach was the KNN method where similarity between user query and training data was calculated.
The final output is given by google search.
It is assumed that words in user query are present in training data.

Results and Discussions

Model	Accuracy
Word Predictor	0.90
Prediction Accuracy	0.85
Average Similarity	0.59
Highest class among top 20	0.75

Summary of Project Outcome

Performance of both the proposed methods gave satisfactory results given the limited data generated for testing and training. There are possible future works that can further improve the models and give better results.

Conclusion and future work

The Deep Learning model uses the top predictions from the Word Predictor model to reduce the search space while identifying and translating each word input from the audio query. The novelty approach was able to perform to a good standard compared to existing approaches

Future work : Increasing the data set, both audio and textual, with a variation in voice such as age, gender, noise level, etc. finding similarity between speech queries can be better done by using spectrograms.

References

1. Multilingual Speech Recognition with a single end-to-end model - Shubham Toshniwal, 15th February 2018

2. Multilingual Text-to-Speech Software Component for Dynamic Language Identification and Voice Switching ,September 2016 Paul Fogarassy,Costin Pribeanu