MULTILINGUAL SPEECH RECOGNITION MODEL FOR RAG WITHOUT TRAINING

SET UP ENVIRONMENT

- First of all we can setup environment to create our project:
  - We can install these python libraries to work on our model

1) Transformers
2) Datasets
3) Sound file
4) Librosa
5) Gradio

LOAD DATASET

- We can load our dataset to train our model and generate effective responses.

- Libri Speech is a corpus of approximately 1000 hours of 16kHz read English speech, we will be using a subset of it for fine-tuning, our approach will involve utilizing Whisper’s extensive multilingual Automatic Speech Recognition (ASR) knowledge acquired during the pre-training phase.

PRE-PROCESS THE DATA

- To pre-process the data to remove noise and extract good voice and generate good audio.
- Noise reduction techniques may be applied to remove unwanted background noise or artifacts from the audio signal.
- Extracting relevant features from the audio signal is a crucial step in audio processing tasks
- Segmenting the audio signal into smaller, manageable chunks can improve the efficiency and accuracy of subsequent processing tasks.
- Data augmentation techniques may be applied to artificially increase the size and diversity of the training dataset.

BUILD THE PIPELINE

- Import pipeline from transformers.
- Selecting model relevant to our project , so in this project we can create multilingual speech recognition using whiper.
- So, we can take existing open source model in hugging face named as **whisper-large-v3** in open ai.
- By using this model we can create multi language speech recognition system. Here over all 99 languages cover.

BUILD A SHARABLE APP WITH GRADIO

- Gradio is a powerful tool and python library for democratizing machine learning by making it easier to share and interact with models, even for users without programming experience.

- It's widely used in research, education, and industry for prototyping, showcasing, and deploying machine learning models with interactive interfaces.

- We can customize the appearance and behavior of the interface using built-in options or by providing custom HTML/CSS styling.

TESTING THE MODEL

- Choose one longer audio file.
- Pre Process the audio file for example to remove noise reduction and voice segmentation.
- We can use librosa for load the test audio files, apply your audio processing pipeline to them, and compare the output against expected results.
- Repeat the testing process until you achieve the desired level of performance and accuracy and save the model.
- Finally we can create a gradio interface and send a audio file into saved model it will give efficient textual data from audio.

CONCLUSION

Multilingual Speech Recognition using the Whisper Large v3 model stands as a testament to the advancements in speech recognition technology and its potential to revolutionize communication and accessibility across linguistic barriers. With its robust functionality, user-friendly interface, and broad language support, this model paves the way for enhanced accessibility, efficiency, and inclusivity in speech recognition applications.

RESOURCE

sunilyadav2713/speech-recognition

THANK YOU