Conversational AI: Speech Recognition (UCS749)

Lab Evaluation I

Submitted By:

Akshat Garg   **102116084   4CS11**

Submitted To:

Dr. Raghav B. Venkataramaiyer

September, 2024

Computer Science and Engineering Department

Thapar Institute of Engineering and Technology

(Deemed to be University), Patiala – 147004

Task Description

The objective of this lab evaluation is to work with the Speech Commands dataset, as detailed in the paper *An Overview of the Speech Commands Dataset*. The tasks include summarizing the paper, performing statistical analysis on the dataset, training a classifier, and fine-tuning it with custom-recorded samples.

Task Breakdown

1. Read and Summarize the Paper
2. Download and Analyze the Dataset
3. Train a Classifier (Google Collab in the Repo)
4. Report Performance Results
5. Create a New Dataset with Custom Samples (google link attached)
6. Fine-tune the Classifier (To be Done)
7. Report the Results

Summary of the Paper (50 Words)

The paper introduces the Speech Commands dataset, which includes 65,000 one-second long utterances of 30 short words by thousands of different people. It aims to provide a benchmark for the training and evaluation of keyword spotting systems, facilitating advancements in voice-activated applications.

Creating a Custom Dataset.

```python
import sounddevice as sd
import numpy as np
import scipy.io.wavfile as wav
import os

def record_audio(duration, sample_rate=8000):
    print("Recording...")
    audio = sd.rec(int(duration * sample_rate), samplerate=sample_rate, channels=1, dtype='int16')
    sd.wait()  # Wait for the recording to finish
    return audio

words = ['backward','bed','bird','cat','dog','down','eight','five','follow','forward',
 'four','go','happy','house','learn','left','marvin','nine','no','off','on','one','right','seven',
 'sheila','six','stop','three','tree','two','up','visual','wow','yes','zero']
# Record a sample
for j in words:
    print(j)
    for kk in range(100):
        continue
    os.mkdir(j)
    for i in range(30):
        audio_sample = record_audio(1)
        wav.write(f"{j}/{i}.wav", 8000, audio_sample)
```

Link to the Dataset in the README file of the git repository.