Lab Eval 1

UCS749: Conversational AI: Speech Processing and Synthesis

B.V. Raghav

[2024-09-11 Wed]

Problem Code: 202425ODD-UCS749-SESS-LE1-0911
Problem Title: Recognise My Voice Commands.

1 Logistics

- Start time: Wed Sep 11 10:00 AM
- End Time: (Extended) Wed Sep 12 11:59 PM
- Submission Form: (Updated) `https://docs.google.com/forms/d/e/1FAIpQLSf9DzoyzW_3kQe2FVqRD7RpjbXmk4HnSun_2LwnWdOggV_q6g/viewform?usp=pp_url&entry.185703634=202425ODD-UCS749-SESS-LE1-0911&entry.1322657816=Recognise+My+Voice+Commands`
- Viva Voce: Will be notified later.

2 Task

Consider the paper: `https://arxiv.org/abs/1804.03209`

1. Read and summarise the paper in about 50 words.
2. Download the dataset in the paper, statistically analyse and describe it, so that it may be useful for posterity. (Include code snippets in your .ipynb file to evidence your analysis.)
3. Train a classifier so that you are able to distinguish the commands in the dataset.
4. Report the performance results using standard benchmarks.

5. Record about 30 samples of each command in your voice and create a new dataset (including a new user id for yourself). You may use a timer on your computer to synchronise.

6. Fine tune your classifier to perform on your voice.

7. Report the results.

3 Deliverables

1. A PDF Report: (as a part of your Git Repo) named <ROLL_{NO}>-report.pdf

2. Assets: Your pretrained classifier model weights and your cleaned and well-formed dataset. This should be a part of your google drive with read access to your instructor <bv.raghav@thapar.edu>

3. A demo notebook: (as a part of your Git Repo), that loads both your model and dataset; and runs to show the results.

4. The demo notebook should verify the assets using a checksum (md5/sha/). This step verifies that the assets have not been tampered with at a later stage.

4 Evaluation

1. Clarity of thought process and presentation.

2. Data processing skills.

3. Model fine tuning/ training skills.

4. Details of progress, as in what were the encountered problems and how were they solved.

5. How adaptable is your pipeline? (as in, how easy is it for me to adapt it for my voice)

6. How scalable is your approach? (as in, how easy is it to scale it to many new voices)

7. Strengths and Shortcomings of your approach.

5 Note

1. This is a test of how fast can we report the performance of a model for a specific task. The best performance is not expected; but a holistic pipeline is.

2. You may improve upon it in future, out of interest; though it wouldnt influence your eval.

6 Struts

The following tutorials may be a good start point; there maybe more on the internet. You are free to choose.

1. `https://colab.research.google.com/github/pytorch/tutorials/blob/gh-pages/_downloads/c64f4bad00653411821adcb75aea9015/speech_command_classification_with_torchaudio_tutorial.ipynb#scrollTo=i0pBRWkcxWrX`

2. `https://colab.research.google.com/github/tensorflow/docs/blob/master/site/en/tutorials/audio/simple_audio.ipynb`