ELEC 378 Final Project - Speech Emotion Classification

Lovingly designed by the TAs

Due: Friday April 21, 11:59PM

1 Introduction

The final project gives you a "hands on" opportunity to apply the techniques you've learned/will learn in both the DSP and Machine Learning portions of this class! The project is intended to not only give you an opportunity to demonstrate your ability to apply the course material, but also to strongly encourage you to explore new concepts and techniques. This year, you are tasked with the classification of speech clips into the emotion of the speaker.

2 Dataset

The dataset consists of sound files across 8 emotions in .wav format. The audio files in the training set will have a label associated with it denoting the emotion it expresses, and the file name. Each audio file is a 3-4 second clip of a vocalization using that emotion.

3 Data Science Process

A good report outlines the entire "data science process," which usually consists of the steps described below. You should follow these steps in order and write about each one in your report.

3.1 Data Exploration

Before doing anything else, you first need to get familiar with your data! You should understand the size and format of the dataset, listen to several individual observations from each category, and experiment with various ways of storing the audio files. This would also be a great time to split your data into training, validation and test sets.

3.2 Feature Extraction

In addition to (or instead of) using raw audio files as input to your classifier, you should engineer features of each file. What are some features of an audio clip? Which ones might be useful for emotion classification? Please provide a description of every feature you extracted during the project.

3.3 Model Selection

You are required to try at least three different model families for this classification task, including one not covered during class. Justify why you chose to try each model. Compare and contrast their performances. Rather than simply stating in your report which model had the highest accuracy, discuss the strengths and weaknesses of each one. Provide visualizations to support your answers.

3.4 Complete Pipeline

Describe the entire pipeline, from preprocessing and feature extraction to model selection and hyperparameter tuning, that you chose to submit. Justify why you think this pipeline will generalize well to the unseen data in the private Kaggle dataset. Include a diagram of your pipeline.

3.5 Conclusions

Summarize your final results on the public dataset, providing visualizations where appropriate. Analyze why you believe you saw the results that you did. Specifically, why did the features you chose help distinguish between emotions? Why did your chosen model outperform others? Lastly, reflect on the process. What did you learn about the data science process that surprised you? What would you do differently in the future?

4 Rules and Expectations

1. You are not allowed to hard code the true labels of the audio files in the test set in any way. Morever, your algorithm should not receive as input the actual labels of the test dataset. We will look through your code for such actions, and your grade will receive a significant reduction if you are found to be in violation. If you don't use any data we haven't provided, you should be fine on this front though.

2. You must use and write in your report on at least THREE different models.

3. At least one of your models must be one we learned in class

4. At least one of your models must be non-neural network based.

5. You are not restricted to using the techniques provided in class. In fact you are strongly encouraged to research and learn other techniques. The caveat is that your report should demonstrate a significant understanding of a technique and its high-level workings, should you choose to use it.

6. You are allowed to use machine learning toolboxes and Deep learning packages etc.

7. You are NOT allowed to use pre-trained models. (Models that have been trained on datasets other than the ones provided to you for this project).

8. You are not allowed to copy another group's code.

9. You are however allowed to reference code (Read: Not copying an entire Git repo) from external resources provided you can explain what the code does. Of course, you must cite it in your report.

10. The code you present in your report must be complete and as you ran it on whatever machine you used.

11. You are allowed to use your own personal computers, any computing resources offered at Rice, and online computing resources (we strongly recommend Google Colab for its free GPUs)

5 Grading

- 10% - Kaggle leaderboard accuracy

- 40% - Technical understanding of new models and methods used

  - This includes the exploration, understanding, and discussion of any new techniques learned outside of class to develop your model(s). This is not just limited to models; it includes any visualizations, dimensionality reductions, feature extraction methods, regularization, ect.

- 50% - General Report Quality

  - How well you discussed the data science process, and why you did what you did. Make sure that there is motivation and justification for every step or procedure you did.

6 Report Submission

Report must be submitted as a PDF, along with any supporting code.