

CS4622 - Machine Learning

Project - Speaker, age, gender and accent recognition using wav2vec base

Dr.R.T.Uthayasanker

August 29, 2023

Project Description

Dataset : AudioMNIST is the dataset used to create the features. Check this link for further details about the dataset [Link](#).

Below structure is general speech-related task classification models

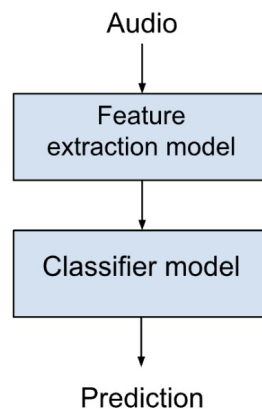


Figure 1: Overview of speech-related task classification model

This project has **two phases**:

1. Phase 1: Individual task - Classification model development and Kaggle competition
2. Phase 2: Group submission - 6-page research paper

Phase 1: Individual task

Wav2vec base is commonly used as a **feature extraction model**. There are **12 transformer layers** in the wav2vec base model. For this project features are extracted from the **last 6 transformer layers** (transformer layer 7 to layer 12). For **each layer separate kaggle competition** is created. **Each train.csv file** provided in the competition contains **layer features** and **corresponding** speaker, age, gender, and accent **labels**.

- Label 1 - Speaker
- Label 2 - Age
- Label 3 - Gender
- Label 4 - Accent

Two (2) kaggle competitions will be **allocated to each person**. **Your task** is to **build classifier models for predicting all 4 labels individually** using features in both training and validation CSV files provided in the competitions. [Find your competition links here [Link](#)]

E.g.

- 1st competition - Layer X
 1. Speaker recognition classifier model using layer X
 2. Age recognition classifier model using layer X
 3. Gender recognition classifier model using layer X
 4. Accent recognition classifier model using layer X
- 2nd competition - Layer Y
 5. Speaker recognition classifier model using layer Y
 6. Age recognition classifier model using layer Y
 7. Gender recognition classifier model using layer Y
 8. Accent recognition classifier model using layer Y

Do data pre-processing, feature engineering, hyper-parameter tuning, dimensionality reduction, cross-validation, and other techniques to improve the classifier accuracy. **Upload the notebook and predicted labels as solutions.csv file** to **Kaggle competition platform** created for this project (More details are provided in the Kaggle competition description, rules sections.)

Phase 2: Group task

Group formation : Maximum 3 people in one group. Only 2 groups can have 2 people.

In your group, the other members would have tried the other two pairs of layers. As a group, your task is to combine all 6 layers and improve the prediction model, and write a 6-page conference paper in IEEE format ([Link](#)).

For the conference paper writing, **do a literature review**, **do ExplainableAI techniques**, and **interpret** the final model. Include your findings from this project and novel ideas during your feature engineering and model development stages in the conference paper. Your paper should be uploaded in easy-chair. The link will be provided later.

The expected content of your conference paper can be found in this [link](#), for your reference.

Evaluation

- **Individual task:**
 - Classifier Model building - **40 marks**
 - * Explainability (Interpreting the label predictions and any cross-relations with labels) - 20 marks
 - * Good practice of ML (right evaluation strategy, ensemble methods, feature engineering, etc.) - 10 marks
 - * Git repository (properly documented) - 5 marks
 - * Coding standard - 5 marks
 - Kaggle Competition Rank - **20 marks**
- **Group submission:** Conference paper - **40 marks**
- We will evaluate your individual task using the ranks from the Kaggle competitions. In addition, the submitted code (notebook / Git repository) will be evaluated based on above mentioned criteria.
- We will evaluate your Group-wise conference paper based on a blind review process considering the quality, findings, interpretations, novelty, etc.

DEADLINES !!

Kaggle competition: 24th September 2023

Paper submission : 8th October 2023