KTH Royal Institute of Technology
Stockholm

School of Electrical Engineering and
Computer Science

Scalable Machine Learning and Deep Learning - ID2223

Project Proposal

Author
Emil Ståhl

Author
Erik Kongpachith

Author
Selemawit Fsha Nguse

December 1, 2021

Project Proposal - ID2223

Emil Ståhl, Erik Kongpachith, and Selemawit Fsha Nguse

December 1, 2021

1 Problem description

In this work, we are going to train a machine learning model to recognize different emotions through spoken sentences. The goal is to recognize at least three basic emotions.

2 Dataset

. To train the model, we make use of the following datasets:

- RAVDESS - This dataset includes around 1500 audio file input from 24 different individuals. 12 male and 12 female where these individuals record short audio clips in 8 different emotions.[1]
- SAVEE - This dataset contains around 500 audio files input from 4 different male individuals where these individuals record short audio clips in 7 different emotions[2]

3 Tools

The tools utilized for this project include:

- TensorFlow
- Keras
- Numpy
- LibROSA
- Matplotlib
- Spark (if required)

For feature extraction we make use of the LibROSA library in Python which is one of the libraries used for audio analysis.

[1] RAVDESS - https://www.kaggle.com/uwrfkaggler/ravdess-emotional-speech-audio

[2] SAVEE - http://kahlan.eps.surrey.ac.uk/savee/Download.html

4 Methodology

The main objective is predicting emotions in sentences from audio samples, one suggestion to the classification problem is to use Convolution Neural Networks (CNNs) for building the model. We are also going to research how models based on multilayer perceptrons and Long Short Term Memory (LSTM) performs in comparison.