Visium Presentation

Kieran Schubert – September 2020

Table of Contents

- Project presentation
- Data Description
- Modelling
- Results
- Demo

Project Description

- Speech Emotion Recognition
- Classification problem
- Create a model
- Create a flask app to serve model (training and prediction)
- Deploy with Docker

Data Description

- Audio inputs (.wav files)
- N=535 data examples
- 7 classes (emotions): Angst, Ekel, Freude, Langweile, Neutral, Trauer, Wut
- 10 speakers
- 10 sentences

Feature Engineering

- Created augmented dataset by
  - Varying length of samples (slower, faster)
  - Varying pitch (0.5 semitones higher/lower)
  - Removing audio «deadspace»
- In a second approach, created 1s samples by randomly sampling the audio files
- Computed MFCCs used as final features
- N = 6955

Modelling

- Start to determine a baseline model
- Fitted different ML models (MNR, kNN, Tree, RF, SVM, Bagged model, Boosted model)
- Fitted different DL models (NN, CNN, LSTM)

Modelling Results

Model	TEST accuracy
MNR	0.688
kNN*	**0.972**
Tree	0.723
RF	0.875
SVM	***0.983***
Bagging	0.686
Boosting	0.779
Bagged kNN	0.957
CNN – Conv1D – no scaling (1)	0.912
CNN – Conv1D – standard scaling (1)	0.774
CNN – Conv1D – MinMax scaling (1)	0.656
CNN – Conv1D – no scaling (2)	0.836
NN	0.746
CNN – Conv2D	0.657
LSTM – Conv2D	0.675

Modelling Results

- Final model is a kNN with hyperparameters determined through grid search Cross-Validation (test acc: 0.969)
- Better performance is certainly possible, but model fine-tuning was not possible for every model due to lack of time

Demo