Scalable Machine Learning and Deep Learning Project Proposal

Yuchen Gao
Weikai Zhou
Group Name: Yuchen & Weikai

December 1, 2021

1 Problem Definition

We would like to collect the voice and generate the corresponding generalized emotion picture for users. For example, if you say something angrily, the computer will generate an angry man picture on the screen.

2 Solution

For the speech emotion recognition, we will conduct operations similar to [1], which includes basically two major parts. The first part is Feature Extractor, where we may do frequency rescaling, normalization, etc. The second part is inputting the extracted features into the LSTM to classify the possible emotion the speaker has. Then the classified emotions will be input into the CNN where the figures of facial expression are labeled. And therefore, we can output the corresponding picture to represent the emotion of the speaker.

3 Dataset

- Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS)[1], which is from the Ryerson University.
- Toronto emotional speech set (TESS)[2], which is from the University of Toronto.
- Facial Expression Detection[3], which is from kaggle.

[1]Source: `https://zenodo.org/record/1188976`
[2]Source: `https://tspace.library.utoronto.ca/handle/1807/24487`
[3]Source: `https://www.kaggle.com/shawon10/facial-expression-detection-cnn/data`

4 Technology Stack

Tensorflow

References

[1] C. Etienne, G. Fidanza, A. Petrovskii, L. Devillers, and B. Schmauch, "CNN+LSTM Architecture for Speech Emotion Recognition with Data Augmentation", *Workshop on Speech, Music and Mind*, 2018.