Hands-On Assignment

Introduction to Deep Learning
800883-B-6

Spring, 2024

Objective

In this assignment, you will explore the application of Deep Learning to analyze Speech. Your mission is to develop a deep learning model that predicts emotional valence from audio signals of spoken language. Remember, the goal here is not just to build a model that works but to understand the why and how behind it. So, put on your data scientist hat, dive into the deep end of deep learning, and let's decode emotions from spoken language together!

Description

This is a group assignment. Groups should be of 3 to 4 people. For this assignment, you will need to implement and train a model and evaluate it on a given test set. You must submit the outputs generated by your model on the test set and a report describing your solution(s) and the code you used. Your model's performance will be compared to that of a baseline model.

This assignment is worth 30% of your course grade. The assignment grade will be based on your work's quality as judged by the instructors based on your **report** and **code**.

Additionally, you will get a bonus based on your ranking on a shared task leaderboard. Specifically:

- if your group ranks first, you will receive a bonus of 2 points on the assignment's grade;
- if your score is no better than a provided baseline, you will receive no bonus;
- for intermediate ranks between the best and the baseline scores, the bonus points will be linearly interpolated. The performance of the baseline solution will be shown on the competition's website.

All the information about the competition submission will be specified in a separate document later.

Passing the assignment is not mandatory to pass the course, but it is highly advisable to hand in your solutions as not doing so implies a 0 on 30% of your final grade (There is no compensation with the exam). There will be no resit for this assignment as passing it is not compulsory, and the

course can be passed without passing the assignment. The exam may include questions that might be easier to answer if you have worked on the shared task.

1 Shared Task – Emotion estimation from Spoken Language using Deep Learning

Human communication is a complex process that goes beyond the words we say. Our voice's tone, pitch, and volume can convey a wealth of information about our emotional state. In this assignment, you will be harnessing the power of deep learning to decode these subtle cues and estimate emotional information from spoken language.

The dataset you will work with for this assignment is a collection of utterances extracted recorded in a controlled environment. This dataset is rich and diverse, featuring utterances from multiple speakers.

Here are the key characteristics of the dataset:

- **Utterances:** The dataset comprises various utterances from different actors. The sampling rate of the audios is 44100.
- **Emotional valence:** Each utterance in the dataset has annotations for emotional valence, a float number between 1 and 5.

Your goal is to build a Deep Learningmodel that is capable of estimating the emotional valence of a recording. Understanding your dataset is the first step towards building a successful model. So, take your time to explore and understand the data.

1.1 Data format

The data set you will receive is a collection of voice recordings. The length of each recording varies, and your model must be capable of handling input of different lengths. The data set will be provided to you in a ZIP file containing a folder with all recordings and annotations in pickle format. Each .pkl file contains a tuple with an utterance as a NumPy array, and the valence values.

Your final model must be evaluated on a given test set that will be provided later. The test set only contains the recordings, so no annotations are given. You have to submit your classification results to the competition to get your model's performance on that set.

2 Deliverables

2.1 Report

A report should be submitted by May 17th, 2024. The report should follow a specific format and include the following:

- **Format:** The report must be correctly formatted and only contain three pages as follows:
    1. A title page where the names and student codes of the group members are listed. Also, the group number must be clearly stated.

2. A single content page (appendices are allowed, but the grade is based on the content page). Please note that anything outside the content page will not be graded except for diagrams on the third page.
3. The third page should only include a visualization of the model proposed and, optionally, other visualizations to help illustrate the architecture. The visualization must be a diagram, not a table. This page must not include text aside from the captions of the diagram.

- The content page must include the following:

1. **Data Preprocessing:** The report should contain a section detailing the data preprocessing steps. It should provide adequate reasoning and arguments for the design decisions made during this process.
2. **Architecture:** The report should describe the chosen architecture. It should argue the decisions made, discuss how suitable and adequate the solution is, and clarify that it is not based on pre-trained models.
3. **Experiments:** The report should briefly describe the experiments conducted, including training, hyperparameter tuning, and optimization processes.
4. **Results:** The report should include a table with the experiment results (including results on the train and test sets) and a table or figure for error analysis. Additionally, the description should give insight into the results obtained.
5. **Discussion and Conclusions:** The report should contain a discussion section that provides clear insights and conclusions that logically derive from the results. This requires valid experimentation and results.

Your report must be a PDF document with a title page, a single page of content, and an additional page with visualizations. Optionally, additional pages for references and appendices can be added. Note that the content page needs to be self-contained, and the appendices should only contain auxiliary material. If any of the above points are in the appendix, they will be considered absent.

2.1.1 Code

Your code should be written in Python and using PyTorch. You do not need to include any data or the weights of your trained model, only the code. The grading of the code will be based on the following points:

1. **Code Submission:** A file with the code should be submitted. The code should be well-documented and easy to understand.
2. **Code Basis:** The code should be based on the course material and should not use pre-trained models (the aim is to learn and understand the process of building and training your own models).
3. **Solution Coherence:** The code should be coherent with the solution proposed in the report. It should effectively implement the proposed architecture and experiments.
4. **Hyperparameters and Settings:** The code should demonstrate that multiple hyperparameters and settings have been investigated. This is an important part of model optimization and should be clearly evident in the code.

5. **Competition Submission:** The results on the test set should have been submitted on the competition website. This will allow for an unbiased evaluation of the model's performance.

Please consider that you can use any approach you consider adequate for pre-processing and feature engineering. You can base your solution on literature as long as you implement it fully yourself without using existing code or pre-trained models.

2.1.2 Submission format

Your report and code should be in a single zip file named with your group ID (e.g., $group_1.zip$) and submitted through the assignment on Canvas.

2.2 Performance

In addition to the report, you must submit your results to a competition. Your submission should be under your group name (e.g., Group 1). **The accounts used should be created with a TiU email; otherwise, it won't be accepted to the competition**. A separate document on the submission to the competition will be provided with additional details later.