Music Mood Classification using CNNs and RNNs

Final Project Fall 2022

Advanced Topics in Data Science (DS-UA 301)

By Mindy Wu & Meghana Kakubal

## Project Overview:

Data
The project consists of a collection of audio files from RAVDESS dataset, which contains actors singing various phrases with specific emotions. The mood category include ‘ emotions = { "neutral": "01", "calm": "02", "happy": "03", "sad": "04", "angry": "05", "fearful": "06", "disgust": "07", "surprised": "08" }.

Preprocessing
Before training the model with CNN and RNN. We load the library librosa where it helps feature extraction in our neural networks. We have CNN and RNN trained on the numpy mean of load librosa array obtained MFCCs. This preprocessing involves converting the audio files into spectrograms and use the spectrograms to calculate the Mel-frequency cepstral coefficients (MFCCs).

Models
There are two models that are trained: the CNN is trained on the MFCCs extracted from spectrograms while the RNN is trained on the sequence of MFCCs.

Evaluation
We evaluate model accuracy based on the performance of model loss and accuracy score on the training and testing set in respect to number of epochs.
The final evaluation of this is tested with five different modern pop songs in five different emotional category: ‘happy’ (Happy), ‘sad’(Special), ‘angry’(Enemy), ‘neural’(Better Days), and ‘lofi’ (Lofi).

## Results:

CNN
Accuracy: 88.67%
Loss: 0.2879

RNN
Accuracy: 73.91%
Loss: 1.041

RNN Confusion matrix on test set:

```
Confusion_matrix:  tf.Tensor(
[[ 0  0  0  0  0  0  0]
 [ 0 20  0  0  1  2  0]
 [ 0  0 40  5  0  0  1]
 [ 0  0  4 35  0  0  7]
 [ 0  0  3  0 29  4 10]
 [ 0  0  0  0  3 32 11]
 [ 0  0  0  2  5  8 31]], sha
```

| | 'true' | CNN | RNN |
|---|---|---|---|
| Happy | Happy | Happy | Happy |
| Lofi | Neutral/Calm | Happy | Fearful |
| Special | Sad | Fearful | Calm |
| Enemy | Angry | Fearful | Happy |
| Better Days | Neutral | Sad | Happy |

From CNN (acc = 88.67%, loss = 0.2879) & RNN (acc = 73.91%, loss = 1.041) performance for music emotion recognition. We can observe that CNN convergence effectively at epochs = 500 while the RNN performed better at early stopping from our audio samples.

CNN model tend to classify pop songs that are true = ‘sad’ or ‘angry’ as ‘fearful’ whereas RNN tend to classify them as ‘happy’ or ‘calm’.

There are still trainings that can be done for model since our model has bias in its:

- CNN tendency in classifying non-happy songs that are neutral or calm into negative emotions.
- RNN tendency in classifying songs into happy (which explains the lower accuracy relative to CNN).

There are areas for improvement in this project. For example, the dataset used in study may not been representative of different music genres and emotions, thus, not generalized well into other datasets. Also, there are features or representations of data that can be further explored and improve performance on.

For future research, we can address these limitations and continuously advance our music emotion recognition.

**GitHub Repository and Code Structure Description:**

- CNN_Music_Mood_Classification.ipynb (CNN Implementation notebook)
- RNN_Music_Mood_Classification.ipynb (RNN Implementation notebook)

**How to Run Code:**

- Please run our notebooks in Google Colab
- You can access the data by adding this data folder to your drive and then modifying the path variable in the notebooks