Song Mood detection

Sai Samarth Taluri
Department Of Computer Science And Engineering
PES University
Bengaluru, India

saisamarth@gmail.com

Sanjana Mahesh
Department Of Computer Science And Engineering
PES University
Bengaluru, India

sanjanamahesh2002@gmail.com

Renita Kurian
Department Of Computer Science And Engineering
PES University
Bengaluru, India

rrenita1206@gmail.com

***Abstract*—We often want to listen to music that best fits our current emotion. A grasp of emotions in songs might be a great help for us to effectively discover music. Automated music mood recognition constitutes an active task in the field of MIR (Music Information Retrieval). In our project, we aim to provide an effective mechanism to classify music into various human emotions based on audio and the metadata. We will be using Mel-frequency cepstral coefficients (MFCCs) extracted from Mel spectrograms as attributes for training our CNN model. In order to extract the MFCC, we will be using Discrete Cosine Transform (DCT) as well as Fast Fourier Transform (FFT). Currently classification of music based on mood is manually done by selecting songs that belong to a particular mood and naming the playlist according to the mood, such as "relaxing". Here we investigate the possibility of assigning such information automatically, without user interaction.**

Keywords—mood, Song Lyrics, CNN, Russell's Circumplex Model, MFCC

I. Introduction

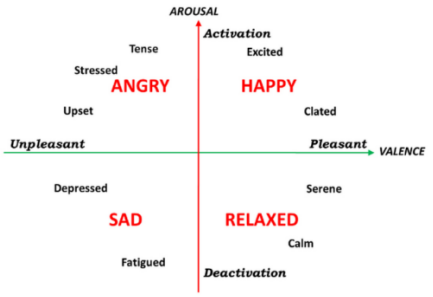
In recent years, automatic playlist generation has been introduced to cope with the problem of the tedious and time consuming manual playlist selection.

Furthermore, browsing the entire music library manually to select songs for the playlist is felt to be difficult by most music listeners. This becomes especially difficult if music collections become prohibitively large, as the user will not know or remember all songs in it.

II. RELATED WORK

A. Data

Songs which are labeled according to Russell's Circumplex Model will be used. Russell used a statistical technique to group the emotion ratings based on positive correlations – which gave rise to 4 major emotions that are "Happy", "Angry" , "Sad" and "Relaxed".



B. Additional Information

The songs in the dataset are of mp3 format which will be converted into a spectrogram using a popular audio analysis library known as "librosa". Spectrogram is a visual representation of the spectrum of the frequencies of a signal as it varies with time.

III. METHOD

A. DataSet

The dataset consists of 1000 songs. They are categorized into 4 classes as per Russel's Model. The dataset also provides a csv file consisting of the top 100 features for the given audio files.

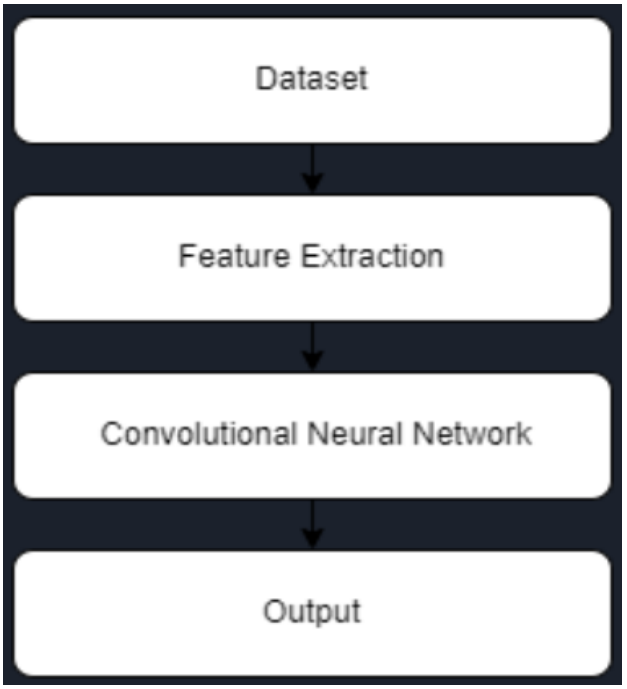
B. Text Pre-processing

- Since our dataset contains no Null values , not much of preprocessing is required
- Feature extraction is the only preprocessing .we have to do
- The library "Librosa" has predefined methods such as Zero crossing rate, chromogram, mel spectrum etc. By trial and error, we came to the conclusion that MFCC (Mel Frequency Cepstrum Coefficients) provide the best results.

MFCC

$$mel(f) = 1127 \ln(1 + \frac{f}{700})$$

C. *Implementation*

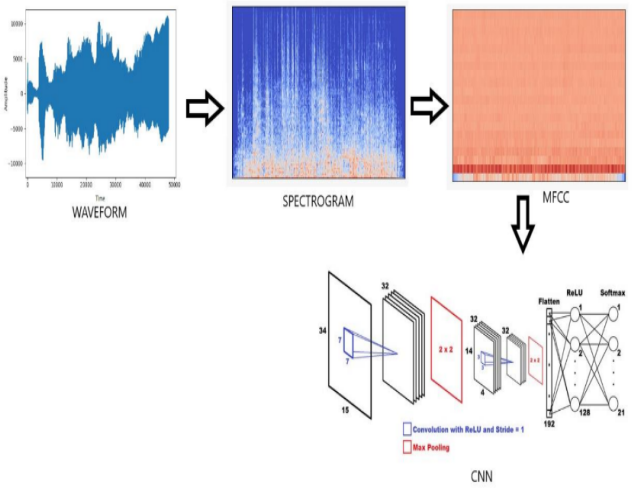


.

Since the dataset already contains labels in the form of q1,q2,q3,q4 depending on the type of the music, and there are no null values, feature extraction is done.

Feature extraction is done using the librosa library which has predefined functions such as MFCC, chromogram, etc. MFCC was seen to give best results and hence it has been used. MFCC feature extraction gives us 40 features from the audio file. These features are then appended to the top 100 features available from the original dataset. Hence, the final feature dataset consists of 140 features for each audio file.

This is then split into a test and train dataset in the ratio 25: 75 and sent to the CNN model. The model consists of 2 dense layers and an output layer. The dense layers have 512, 256 and 4 units. We have also used dropout layers to reduce overfitting. 'Relu' activation function has been used along with softmax for the output layer. The optimizer used is adam. Categorical cross entropy loss function is used as it performs well for multiclass classification.




```
Model: "sequential_1"
_________________________________________________________________
 Layer (type)                 Output Shape              Param #
=================================================================
 dense_3 (Dense)              (None, 512)               72192

 activation_3 (Activation)    (None, 512)               0

 dropout_2 (Dropout)          (None, 512)               0

 dense_4 (Dense)              (None, 256)               131328

 activation_4 (Activation)    (None, 256)               0

 dropout_3 (Dropout)          (None, 256)               0

 dense_5 (Dense)              (None, 4)                 1028

 activation_5 (Activation)    (None, 4)                 0

=================================================================
Total params: 204,548
Trainable params: 204,548
Non-trainable params: 0
_________________________________________________________________
```

The model then returns the output in the form of text. The testing data is classified either as happy, angry ,sad or relaxed.

IV. RESULTS

We have tested using multiple classification models like SVM, XGBoost, KNN, Naive Bayes and CNN and CNN came on top with around 73% accuracy.

```
Test accuracy of CNN: 72.89%
Test accuracy of XGBoost Model is: 57.78%
Test accuracy of SVM Model is: 68.89%
Test accuracy of KNN Model is: 64.00%
Test accuracy of Naive Bayes Model is: 68.00%
```

.

Acknowledgment

We would like to thank our instructor from the department of computer science at PES University, India. In addition, we would like to thank the university for providing us with an opportunity to complete this project.

Abbreviations Used

MFCC - Mel Frequency Cepstrum Coefficient
CNN - Convolutional Neural Network

References

Some of the papers we used for reference are mentioned below

[1] MOOD CLASSIFICATION USING LISTENING DATA By Filip Korzeniowski, Oriol

[2] TRANSFER LEARNING FOR MUSIC CLASSIFICATION AND REGRESSION

[3] EMOTION BASED SEGMENTATION OF MUSICAL AUDIO By Anna Aljanak