Music Emotion Recognition using Deep Neural Networks

T.H. Bathigama[*] and M.K.S. Madushika

Department of Computer Science, Faculty of Science, University of Ruhuna
*Corresponding Author E-mail: bathigama9935@usci.ruh.ac.lk, TP: +9470 244 3757

Emotion is an integral part of music and a complex aspect of music that is not easily understood by machines. The emotional aspect of music is further complicated by the fact that it is a subjective experience that cannot be easily conveyed to machine. Although it is a complex problem, some progress has been made in this area suggesting that it might be feasible to develop computation models that can be used in real-world applications. Real-world applications of music emotion recognition systems range from entertainment to healthcare. In this paper we introduce a deep learning model that recognizes emotion in music from the audio signal. 1d and 2d convolution layers with different kernel sizes have been tested. Adaptive pooling layers have also been used to extract a fixed feature representation for the dense layers. We have also used trainable spectrogram extractors to learn different representations of the audio. To address the lack of data for the task of music emotion recognition we have also used the latest trends in audio data augmentation and converted it for music data. Till now we have been able to achieve an accuracy of about 0.92 for the PMEmo dataset and about 0.6 F-1 score from using the raw audio signal and 1D convolution layers to extract features. Preliminary experiments show that using 1d convolutions with the combination of learnable spectrograms performs satisfactorily. Further experiments are to be conducted using different combinations of raw audio and calculated features. Different model architectures using recurrent networks are also to be tested considering that audio has temporal relationship between each unit of time. Finally, the work done in this study is mainly to explore the high dimensional feature space of raw audio to extract features which can contribute to the recognition of emotion in music using automated methods such as convolution and recurrent layers.

***Keywords*:** Music Emotion Recognition; Deep Neural Networks; Music Data Augmentation; Arousal and Valence Prediction