

MusicGuru: Multi-task and Multimodal Representation Learning for Music Applications

Project Abstract

by

CMPE 295B

Arpitha Gurumurthy (arpitha.gurumurthy@sjsu.edu)]
Bhuvana Gopala Krishna Basapur (bhuvanagopalakrishna.basapur@sjsu.edu)
Mayuri Lalwani (mayuri.lalwani@sjsu.edu)
Somya Mishra (somya.mishra@sjsu.edu)

Project Advisor

Prof. Vijay Eranti

December 2021

ABSTRACT

MusicGuru: Multi-task and Multi-modal Representation Learning for Music Applications

By

Bhuvana Gopala Krishna Basapur, Arpitha Gurumurthy, Mayuri Lalwani, Somya Mishra

The way in which we perceive the world is by using our sense of sight, hearing and smell, which is multimodal in nature. Research shows that multimodal learning helps in better understanding and retention of information in humans. Multimodality in AI refers to training models by combining multiple modes of input such as visual, audio and textual representations. There are correlations between them which further help to uncover useful patterns in the data that may not be obvious when using a unimodal approach. Similarly, a multi-task learning approach in AI trains a model to perform different tasks. This helps in generalizing the model as it learns the commonalities between these tasks and helps in achieving better performance for all the tasks. Introduction of multimodality and multi-task applications in AI will train models to be more ‘human-like’ and help in moving one step closer to ‘self-improving AI’.

Since multimodal learning involves engaging multiple senses in the processing of information, it is considered more effective than single-modality learning. The concept of music is also multimodal. Using only audio to express music would be very restrictive. As discussed in the literature search section of this workbook, it is seen that most of the previous works in the field of music understanding focused on using only audio tracks or symbolic music as input. Text based approaches have also been employed but research using multimodal input is still limited. Music related tasks would hugely benefit from multimodal learning approaches as it can learn the fine grain details present in the language of music. Musicians and composers have mostly lacked a similar tool for exploring and mixing musical ideas, but the objective of this application is to change this. MusicGuru, is an application that would help creating palettes for blending and exploring musical notes.

The goal of this project is multimodal music understanding using intelligent DL techniques. To take it one step further, we aim at building a model that can jointly handle multiple music related downstream tasks to solve issues concerning extensive data and computational requirements. The concept of multimodality and multitasking in the field of music help in better understanding of music structure in terms of its melody, rhythm, beat and chord. This in turn results in better performance for tasks like melody completion, changing the pitch while keeping the rhythm constant, changing the rhythm while keeping the pitch constant, changing the melody of the song with the existing chord progression and changing the chords of the song with the existing melody progression as the model has a provision of visualizing, hearing and finally analysing the same input.