CSC413 Project Proposal

Tomasz Cieslak (1005361948, cieslak4)
Mina Makar (1005087282, makarmin)
Daren Liang (1005402137, liangd10)
George Lewis (1005013183, lewisge2)

Task of the model

Music has been historically categorized by humans into several genres, such as rock, blues, classical, and hip-hop. These genres are difficult to define precisely, but the songs of each genre always share a number of similarities. These genres are easiest to define by the music that is already considered to belong to them. This makes this task difficult to encode into a traditional algorithm, and so we believe that a neural network is uniquely suited to the problem, as it can learn common features of genres that are hard to define rigorously.

We will create a model that analyzes segments of digital music recordings to predict the genre, mood, theme, and instruments of the sample. The model will rely on pattern recognition based on the sequence of features extracted from different time sequences of the musical input. To make our model competitive compared to other implementations, we will also utilize various attention-based mechanisms to optimize the performance and predictive accuracy.

Model breakdown

Audio is a sequence of data usually represented using pulse-code modulation[1] which creates a time-series of samples by quantizing analog audio at regular intervals. The time-sequence nature of audio data means that it is well-suited to a recurrent neural network (RNN) architecture, as a sequence of features will be extracted from different time sequences, which will capture the temporal dynamics of the input.

Since each song carries an overarching theme and mood, we can use LSTM or GRU to retain information and find long term dependencies throughout the time sequences to categorize mood and theme, and possibly the genre of the musical input.

Available datasets to train the model

We will be using the MTG-Jamendo dataset[2] which is a relatively new music dataset created in 2019 and contains more than 55,000 tracks with 195 tags. All of the tracks in the dataset can be found on Freesound.org, a creative-commons licensed music library.

The relevant fields in the dataset that we are going to use are path and tags (assuming the music files are already downloaded). Here is a small preview of the dataset:

Track ID	Artist ID	Album ID	Path	Duration (s)	Tags
0000382	000020	000046	82/382.mp3	211.1	classical, gospel, voice
0340682	350154	045019	82/340682.mp3	250.2	chillout, downtempo, electronic, lounge, nujazz, symphonic, triphop
0922855	419555	109326	55/922855.mp3	159.7	ambient, grunge, rock, soundtrack, bass, drum, electric guitar

A subsection of the audio track utilized during training may vary in duration and tags

The 195 tags in the MTG-Jamendo dataset consist of 95 genres, 41 instruments, and 59 mood/themes. The 5 most popular genres are electronic, soundtrack, pop, ambient, and rock. The 5 most popular instruments are piano, bass, synthesizer, drum, and electric guitar. The 5 most popular mood/themes are happy, melodic, dark, relaxing, and energetic. All tags in the dataset are guaranteed to have at least 100 tracks assigned to each.

The dataset may include tracks that are very short, so it is a good idea to ensure only tracks that are longer than a certain duration get included in the dataset used to train our model. A good cut-off duration is 30 seconds and the filtered dataset will contain 55,701 tracks in total.

Ethical implications

We believe this model to be relatively free of any ethical implications, as all the music in the MTG-Jamendo dataset is available under a Creative Commons Attribution Non-Commercial Share Alike license. If we plan to source any additional music we will ensure to verify its license.

We see our model as helpful, as it can be used by other audio/music hosting applications when attempting to sort their data by categorizing their music libraries. Moreover, it could highlight similarities between genres, and may uncover certain patterns between them, which could augment research in the musical space.

Team contribution breakdown

- Tomasz Cieslak - Data integration, training model (locally), model implementation
- Mina Makar - Data integration, model implementation, final readme
- Daren Liang - Data sourcing, formatting and integration, model implementation
- George Lewis - Model research, data sourcing / integration, model implementation

References

1. https://en.wikipedia.org/wiki/Pulse-code_modulation
2. https://zenodo.org/record/3826813