Music Genre Recognition Using Machine Learning

Joseph Byrne - 2017118092

***Abstract*—Music genre classification is an important task with many real-world applications. Companies such as Spotify and Shazam use music data in many ways to provide services to their users. Machine learning can be used to increase the efficiency if music databases that categorize songs into different genres. Its also useful for categorizing music for use in recommendation systems and automatic playlists generators that service providers like Spotify use. In this report, the automatic classification of audio signals into musical genres is explored.**

***Index Terms*— Machine Learning, Music Classification, Audio, Audio Classification**

I. Introduction

MUSIC genres are labels used for categorizing and describing the vast world of music. Every day more and more music is being released and as the amount of music grows, so does the need for accurate meta-data for efficient database management and search functions. The ability to quickly and easily classify any song into a music genre in any given playlist or library us a very useful function for any music service provider. An implementation of a Rectified Linear Activation Function is used to create a model. The GTZAN dataset's .wav files are converted to Mel-frequency cepstral coefficients (MFCCs) and inputted into the model, using each .wav file's respective folder as a label.

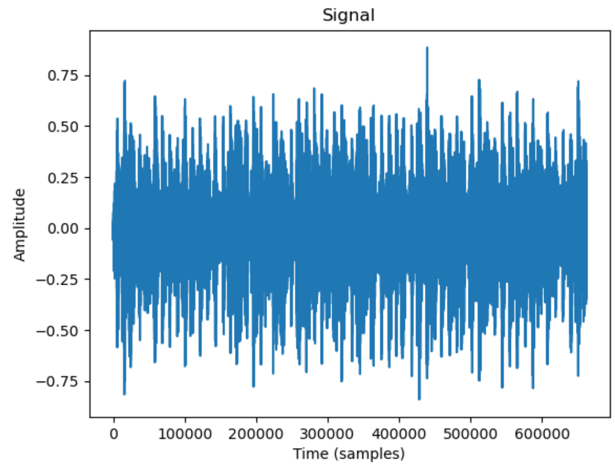
There are two main issues with this problem:

- Music genres are loosely defined
- Its difficult to find the exact variations in songs that should be inputted into the model

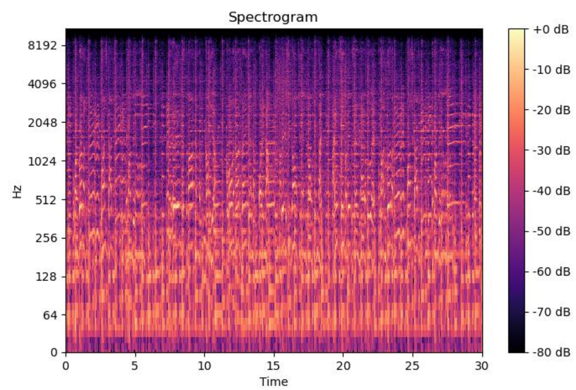
The first problem we can't control as it's a natural aspect of music and is also partially subjective. The second problem has a lot of research on the topic, in the field of Music Information Retrieval, which main concern is with getting the most important information from audio signals.

II. The Dataset

The GTZAN Genre Collection is a famous dataset that was used for a well-known paper in genre classification called "Musical genre classification of audio signals" by G. Tzanetakis and P. Cook. The files were collected from 2000 to 2001 from of sources including personal CDs, radio, microphone recordings, in order to represent a variety of recording conditions. The dataset consists of 1000 audio tracks each 30 seconds long. It contains 10 genres, each represented by 100 tracks. The tracks are all 22050Hz Mono 16-bit audio files in .wav format and each is 30 seconds long.



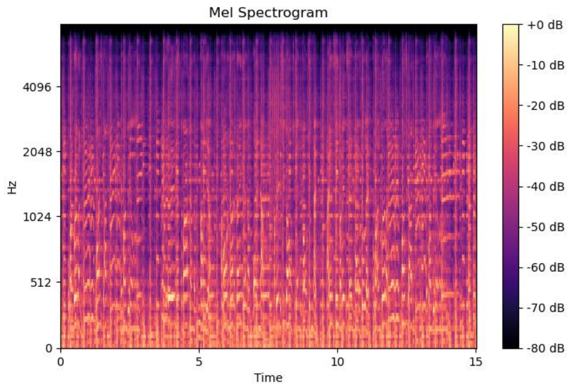
This image is a visual representation of the first blues song in the dataset (blues.00000.wav) in terms of amplitude over time. Compared to other genres its possible to see a difference just based on the this. However, it can be even easier for a machine to tell the difference just based on an image like this, but this was not the input used for the model.



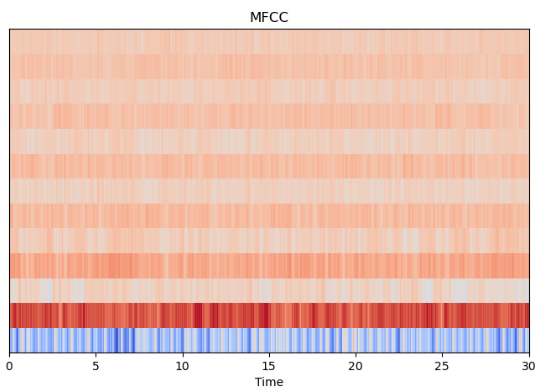
A spectrogram, as seen above, is a visual representation of frequencies of an audio signal as it varies over time. The spectrogram in the image above is also of the blues.00000.wav file. In this spectrogram representation plot, one axis represents the time and the other represents the frequencies. The colours

.

represent the observed frequencies at a particular time. Smaller frequencies are brighter.



The mel spectrogram is another visual representation of an audio signal where frequencies are converted to the mel scale, and represents how frequencies change over time and can be seen in the image above (using the same audio file as the others). Since humans don't perceive frequency on a linear scale we map the frequencies to a mel scale to measure the pitch. This makes it so that equal distances in pitch sound equally distant to the human ear.



A popular audio feature extraction method is the Mel-frequency cepstral coefficients (MFCC) which will be used as the input for the model as it provides us enough frequency channels to analyze the audio.

III. Preprocessing the Data

As we know, to train machine learning models we need a lot of data, so since we only have 100 data points in the form of tracks for each genre which it isn't a lot. So we need to divide each track into different segments so that we have more input data. A dictionary of key pair values of MFCC files and labels was made. So rather than saving the inputs as tracks, the inputs were saved as individual segments. The data is stored in a json file which is used as the input for the model

IV. The Model

20% of the data was used as the test set and 80% was used to train the model. A sequential model with an input layer, three hidden layers and an output layer was used. Since we are working with a simple multi-layer perception, dense layers were chosen to be used.

A summary of the model can be seen here:

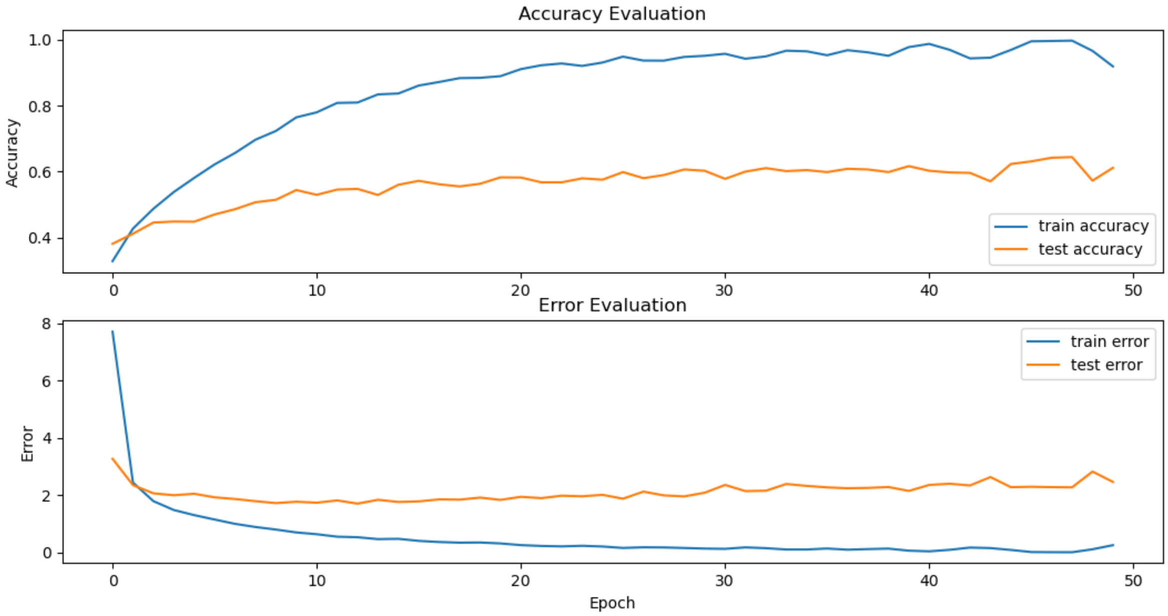
Model: "sequential"

Layer (type)	Output Shape	Param #
flatten (Flatten)	(None, 1690)	0
dense (Dense)	(None, 512)	865792
dense_1 (Dense)	(None, 256)	131328
dense_2 (Dense)	(None, 64)	16448
dense_3 (Dense)	(None, 10)	650

Total params: 1,014,218
Trainable params: 1,014,218
Non-trainable params: 0

Train on 7996 samples, validate on 2000 samples
RELU was used for each later.

V. Conclusion



There is a big difference between the accuracy on the test set and the accuracy on the training set. The model preforms great on the training data but not so well on the test data. It seems like the model is overfitting. If this was done again, some overfitting solutions such as dropping nodes should be tried or stopping training early should be used. Implementing regularization could also be useful and applying transformations to the data such as pitch shifting, time stretching and adding background noise.

References

[1] Musical Genre Classification of Audio Signals
George Tzanetakis, Student Member, IEEE, and Perry Cook, Member, IEEE

[2] Music Genre Classification - http://cs229.stanford.edu/proj2018/report/21.pdf

[3] Musical Genre Classification with Convolutional Neural Networks - https://towardsdatascience.com/musical-genre-classification-with-convolutional-neural-networks-ff04f9601a74

[4] Music Genre Classification using Machine Learning Techniques - https://arxiv.org/abs/1804.01149