Ukrainian Catholic University

Faculty of Applied Sciences

Music Genre Classification

Basics of Signals and Systems project report

Author:

Volodymyr Savchuk

23 December 2021

1. Introduction

Automatic musical genre classification has been a quite popular area of research in recent years. A music genre is one of the most important labels of the songs. It distinguishes the songs and is the main category that listeners pay attention to and consider. Also, with the rapid development of technology, audio streaming services, sound recording techniques, the number of audio resources, the number of songs show colossal growth. Users want to have an easy label for music classification in the way of music genre, so we need somebody or something to define the genre of the music. Manual music genre classification is time-consuming and expensive because professionals should do it with a professional background in music. In order to classify a massive amount of songs, we will use a lot of person-hours and money, and the classification still will be subjective as there is no unified classification system. The automatic musical genre classification system could solve the problem of manual classification cheaply and rapidly. Finally, my research aims to analyze the current approaches to classify the music genre, try to implement my own, and compare results.

2. Literature review

An automatic genre classification problem algorithm could be helpful for many applications. For example, it could also help music streaming platforms like Spotify and Apple Music.

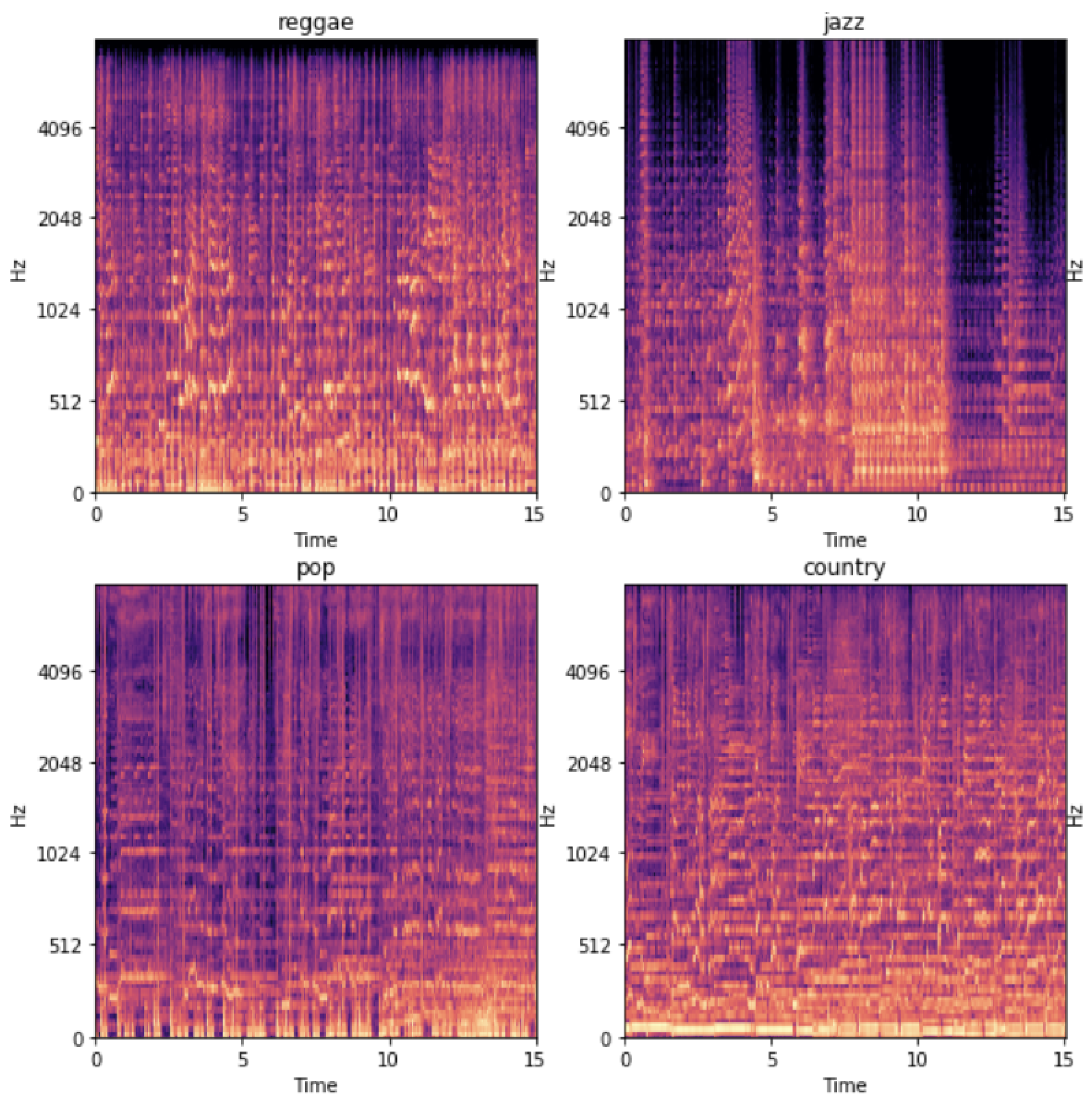
The first work is the one of Tzanetakis and Cook [1]. They used three different feature sets as a way to represent timbral texture, rhythmic and pitch content. Short-time Fourier Transform (STFT), Mel-frequency Cepstral Coefficients (MFCCs), Wavelet Transform (WT). With these features, they trained such classifiers as Gaussian Mixture Model and *k*-Nearest Neighbor, simple Gaussian, using real-world audio collections. The achieved result is 61%.

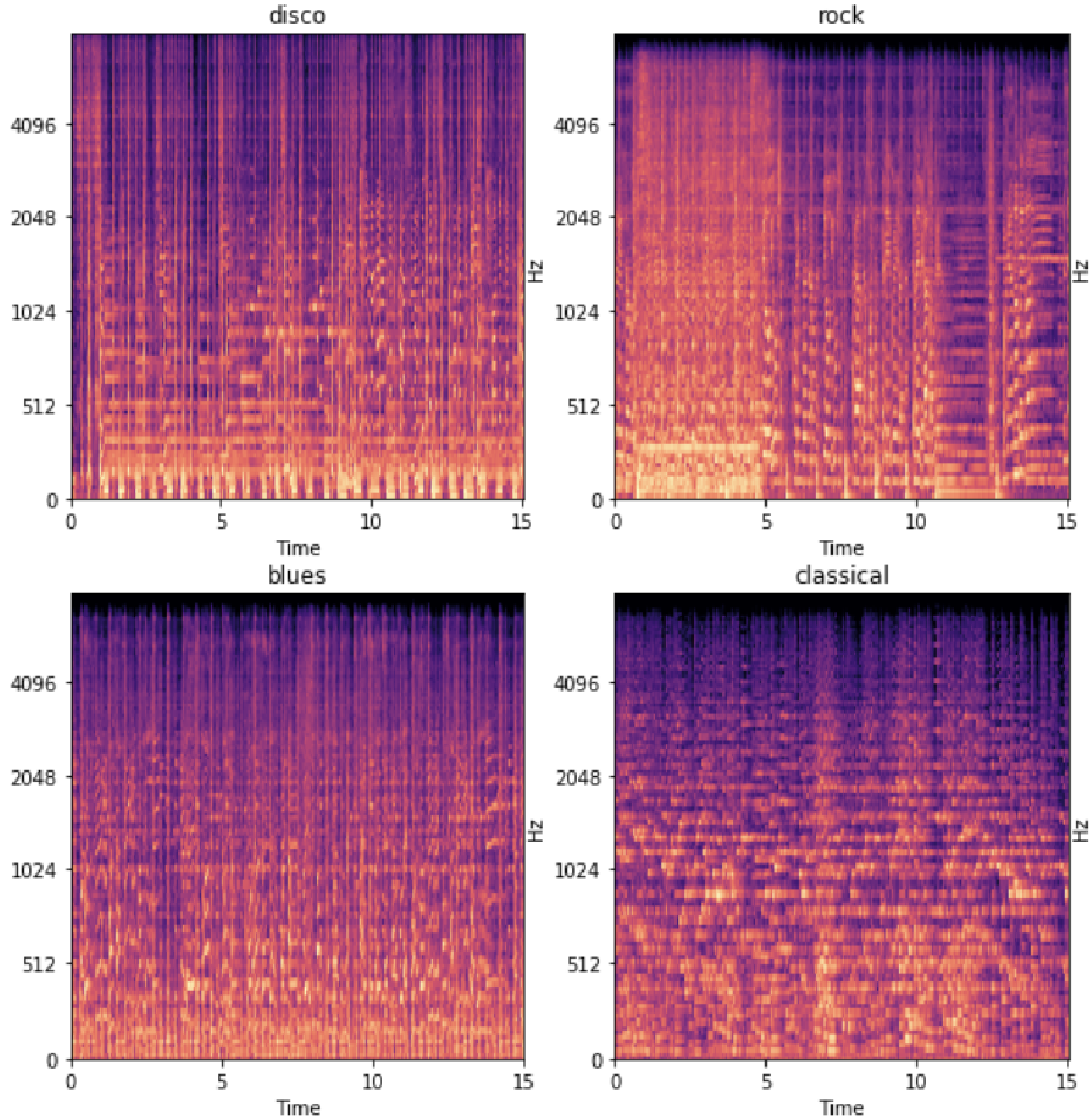
The second work is done by Panagakis and Kotropoulos [2]. They proposed a musical genre classification system that uses as features 2D auditory temporal modulations. The results are astonishing, namely 91% accuracy on the GTZAN dataset[3].

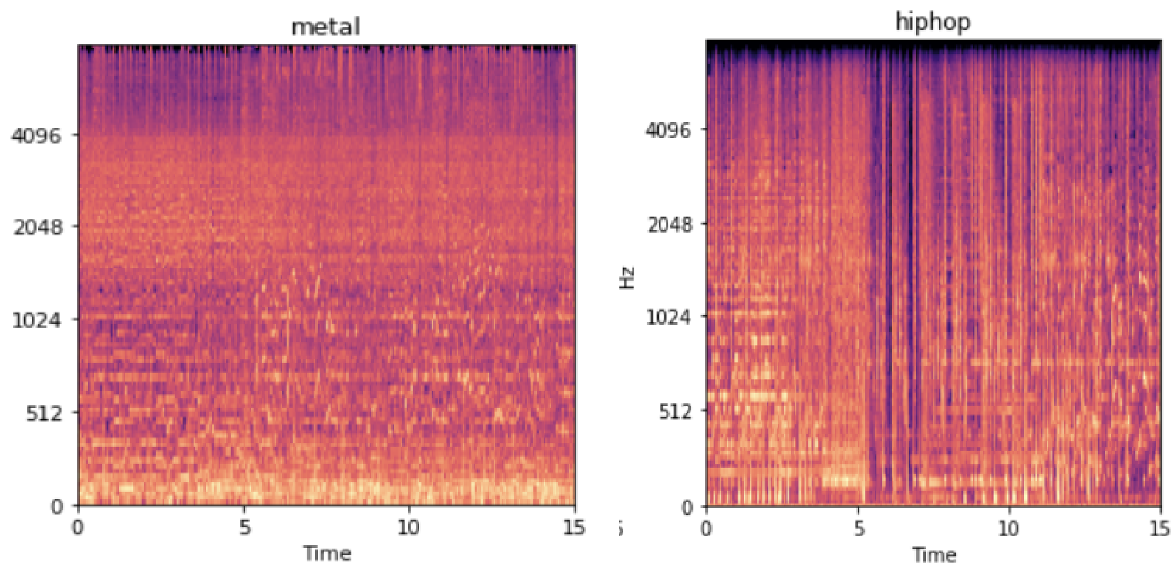
The third work is Silla et al. [4]. They used features that were selected from different parts of the song. *K* Nearest-Neighbors, Naive-Bayes, SVMs, decision trees, and Multilayer Perceptron Neural Networks were trained. The best accuracy was 65.06%.

3. EDA

The dataset, which will be used for the research is GTZAN[3]. This dataset consists of 1000 audio tracks, each 30 seconds long. There are ten genres, each containing 100 tracks, all 22050Hz Mono 16-bit audio files in .wav format. The genres are blues, classical, country, disco, hip-hop, jazz, metal, pop, reggae, rock. Let us look at Mel Spectrograms of ten tracks with ten different genres.




disco
rock
blues
classical
4096
2048
1024
512
0
Hz
0
5
10
15
Time



These plots show us that there are some patterns in Mel spectrograms, which each genre possesses so that they could become great features for the classification model.

4. Selected approach

The main task is somehow to classify the audio file, and the solution is to make a model, which will make this classification. The next is to choose the input features of the model. Here the theory of signals and systems comes in handy. I will use the notion of Mel spectrogram, which in simple words, is a visual representation of an audio signal. In other words, it shows how the spectrum of the signal frequencies changes over time. Why is it better in our case than the usual spectrogram? The reason is in how humans perceive frequencies. For example, the sounds of frequencies 100 Hz and 200 Hz will sound much different from those of 1100 Hz and 1200 Hz. Here Mel spectrogram comes in handy, as it uses the Mel

Scale instead of Frequency on the y-axis and the Decibel Scale instead of Amplitude to represent colors, where the Mel Scale is a scale of pitches.

The Mel spectrograms are represented in the form of NumPy array 128×660, but we need to add one dimension representing a single color channel. Also, the Mel spectrograms have to be scaled to the values between 0 and 1 for making calculations quicker.

Then the CNN model training and validation were performed, trying different model architectures.

5. Results & Conclusions

The best CNN model has a test accuracy score of 60 % (based on a test accuracy score). Also, this model has a training accuracy score of 64%, which means the model almost hasn't overfit. However, the results are still weighty. Looking at a confusion matrix, we can see that model mostly makes mistakes by confusing genres humans also confuse. For example, reggae vs. hiphop, blues vs. jazz.

Plans and suggestions for improvements could be to go deeper into Feature Engineering. Other components from audio samples could be used as features: Chroma(harmonic and melodic characteristics), Zero Crossing Rate, Mel-Frequency Cepstral Coefficients(MFCCs), Spectral centroid, Spectral Rolloff. All except the first one Chroma were extracted from the audio files and I plan to use it as features in the future.

Literature:

[1] -
https://www.scopus.com/record/display.uri?eid=2-s2.0-0036648502&origin=inward&txGid=eae4f4d07cc6ea42c19fa172a6e781c2

[2] -
https://scholar.google.com/scholar?q=Panagakis%20Y,%20Kotropoulos%20C,%20Arce%20GR.%20Music%20genre%20classification%20via%20sparse%20representations%20of%20auditory%20temporal%20modulations.%20In:%2017th%20European%20signal%20processing%20conference%20;%202009.

[3] - http://marsyas.info/downloads/datasets.html

[4] -

https://scholar.google.com/scholar_lookup?title=A%20machine%20learning%20approach%20to%20automatic%20music%20genre%20classification&publication_year=2008&author=C.N.%20Silla%20Jr.&author=A.L.%20Koerich&author=C.A.A.%20Kaestner