



Music Tagging

06-08-2018

Mentors:-

Mansi Sampat

Vasu Eranki

Team Members:-

Dharmik Bhatt

Insiyah Hajoori

Gulshan Rana

Yash Mithaiwala

Swastik Bhattacharya

Drishti Mini-project

Aim

To develop a model using machine learning algorithms which will predict the genre of a song by itself.

Goals

1. To develop a machine learning code and test it on a dataset.
2. To classify the songs in two genres.
3. To develop a gui.

Abstract

Music tagging or music tagger is a software which can classify songs into their respective genre. This project included many subparts which are further mentioned in this document.

In this mini-project, we have used logistic regression(binary classification) but the same or more better results can be achieved by SVM(Support Vector Machine) and Neural Networks.

This kinds of software are mainly used in apps like saavn,spotify,gaana,etc where music of same genre are classified together.

What kind of data is being dealt with?

To understand this, one must need to know a few basics of music theory. Any music (song, in this case), consists of two components: notes and beats. Manipulating these two gives rise to many different songs, which can grouped into different genres.

Our brain, upon listening to some music, conceives a pattern, which enables us to recognize or identify the genre of music. This pattern consists of different arrangement of notes and beats. So, we are using machine learning algorithms to train the machine recognise a pattern. This requires some features, which are:

- Beats
- Tempo
- Chroma Short Term Fourier Transform
- Root Mean Square Error
- Spectral Centroid
- Spectral Bandwidth

- Spectral Roll-off
- Zero-Crossing Rate
- Mel Frequency Cepstral Coefficients

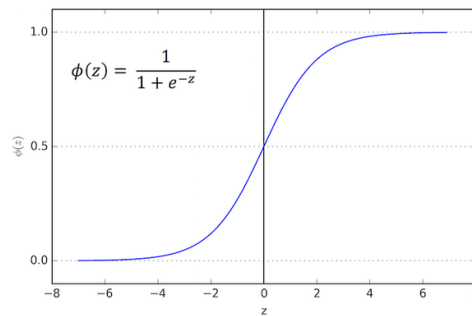
But first, let's see how our brain conceives different features of music.

Tempo is the rate of number of beats appearing in a song. This gives us a notion of speed. The other thing that matters in classification of a song into genres is the arrangement of notes. Every song of a particular genre has a particular pattern of notes appearing. For e.g. in blues, minor notes appear suddenly in a flow of major notes, giving that melancholic feeling. Furthermore, the instruments being played is also necessary for classification. For e.g. saxophone is predominant in jazz; in pop, guitars; in classical, cello; etc. These features can be implemented into machine learning by the above mentioned features, as these give a descriptive idea about the qualities of the sound signal in the music.

This project includes...

1. Dataset extraction
 - a. A music genre is a conventional category that identifies pieces of music as belonging to a shared tradition or set of conventions. It is to be distinguished from musical form and musical style. The features extracted from these waves can help the machine distinguish between them.
 - b. The features in this dataset are extracted from the *GTZAN genre collection* provided at the site of *Marsyas (Music Analysis, Retrieval and Synthesis for Audio Signals)*. It is an open source software framework for audio processing with specific emphasis on Music Information Retrieval applications.
 - c. It consists of 1000 audio tracks each 30 seconds long. It contains 10 genres, each represented by 100 tracks. The tracks are all 22050 Hz Mono 16-bit audio files in .wav format.
 - d. Features are extracted using *libROSA* library. LibROSA is a python package for music and audio analysis. It provides the building blocks necessary to create music information retrieval systems. The extracted features were saved in a csv file.
2. Logistic regression:-
 - a. Logistic regression is used used to classify data into groups(binary classification i.e two groups) .
 - b. Logistic function is a sigmoid function which gives values in between 0 and 1.
 - c. Logistic regression includes further many subparts:-
 - i. Hypothesis representation- Hypothesis is a function which we will feed into machine and the machine will process the parameters using that function. Hypothesis for logistic regress is a sigmoidal function.

- ii. Cost function- Cost function is basically the difference between the machine predicted value and the actual value.
- iii. Decision boundary- This concept includes rounding off of the predicted result to either 1 or 0. For eg, if the programmer sets the decision boundary to 0.6, then all the predicted values below 0.6 will be rounded off to 0 and those above 0.6 will be rounded off to 1. By changing the value of decision boundary programmer can increase the efficiency of the program..



- d. The above image represents the sigmoid function and its equation.
 - e. Logistic regression can also be used to classify data into more than two classes but it's discussion is beyond the scope of this documentation.
3. Gradient descent
- a. Gradient descent is a first-order iterative optimization algorithm for finding the minimum of a function.
 - b. To find a local minimum of a function using gradient descent, one takes steps proportional to the negative of the gradient (or approximate gradient) of the function at the current point.

Problems faced

1. Our dataset was in csv format, we needed to extract our data in the form of array which must be compiled with our proposed library (in our case it was NumPy). Earlier we used csv reader to read the dataset, but dataset was extracted as an object datatype. so afterwards we used genfromtxt inbuilt NumPy which created the dataset array.
2. Matrix multiplication was creating a problem because dataset was containing data of different datatypes.

Proposed Approach

1. Since the music genre classification was to be done by machine learning, we needed to train our machine to differentiate the genre on the basis of various features.
2. Features that we included in our project were beats,tempo, spectral centroid, spectral roll-off, Chroma frequencies, zero crossing rate, mfcc's etc.
3. Once the feature vectors were obtained, we trained different classifiers on the training set of feature vector. There are many classifiers for this purpose such as logistic regression,svm(support vector machine),K-nearest Neighbours,etc.Neural network could also be used for this purpose. The one we used here is logistic classifier since it worked well for classifying two genres with greater accuracy.

Result

We started by classifying songs into 2 genres (pop, classical) using logistic regression and iterate the parameters' values. It gave an accuracy of around 90-95% over testing data.

Future aspects

This project can also be accomplished by using neural networks,SVM and Logistic regression(one vs all). All of the above mentioned techniques can be used to classify audio files into more than two genres.

References

1. <http://marsyas.info/downloads/datasets.html>
2. <http://librosa.github.io/librosa/>
3. <https://www.kaggle.com/insiyeah/musicfeatures>