International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 08 | Aug 2019 www.irjet.net p-ISSN: 2395-0072

Audio Genre Classification using Neural Networks

Gandharva Deshpande[1], Sushain Bhat[2]

[1,2] *Student, Dept. of Computer Engineering, St. Francis Institute of Technology, Mumbai, Maharashtra, India*

Abstract – *Audio Classification is a fundamental problem in the field of Audio processing. One of the modern applications of such classification is the Audio Genre Classification. With the music industry soaring high as a form of entertainment, the classification of the genres of music became an important aspect. As a solution to this problem of classification, we built a Neural network model to predict the genre of a music file. The model essentially uses a Neural network architecture to classify the music file into 10 different genres based on various parameters of the audio clip. The model is trained using audio files and then further tested using a set of audio files.*

Key Words: **Neural Networks, Audio Dataset, Genre Classification, Audio files, Classifier model**

1. INTRODUCTION

Audio Genre Classification is a task that essentially requires human intelligence to some level. Determining a genre of music simply by listening to it may require some level of intelligence and experience on the human level too. To ease things down for the music enthusiasts, and help them in the process of determining the genre of a music piece, the model classifies the audio file into the most suitable genre. The objective of the model is predicting the genre of the music file using neural networks with considerable accuracy and respectable speed. This concept paves way for the rising machine learning applications in the field of music industry.

1.1 Dataset Description

The Dataset for the model is an audio file dataset consisting 1000 audio files of 10 different genres. The file format of the audio files was .mp3. A total of 800 audio files were used for training purpose while the rest 200 files were utilized for testing. The dataset comprised of varied audio files however only a duration of 30 seconds is considered for feature extraction purposes.

1.2 Model Selection

When it comes to dealing with audio files or image files neural network models are the Go-to models in machine learning applications. With an aim of achieving higher accuracy a 5 layered neural network model was developed to attain efficiency. The model consists of the following:

4 layers with RELU activation function, Output layer with SoftMax activation function

2. Hardware and Software Requirements

Requirements are the minimal configurations of a device and software required for the model to work properly and efficiently.

2.1 Hardware requirements

Graphics Processing Unit (GPU).

Intel Core i3 processor or above

2.2 Software requirements

Windows 7 or above / Linux.

Python 2.7 or above.

Jupyter Notebook.

3. METHODOLOGY

The classifier model aims at providing respectable accuracy and speed in successful classification and prediction of genre based on the input audio file. Initially, training data set is provided to the model. The model learns to classify genres and predict the genre for the input file with the help of training. A Neural network model is used for the same. The node structure is developed based on the various features that help determine genre of a song or a music file. The essential features (features that correlate the maximum with the genre of a song) are utilized to rightly classify the audio track. The workflow of the system is given below:

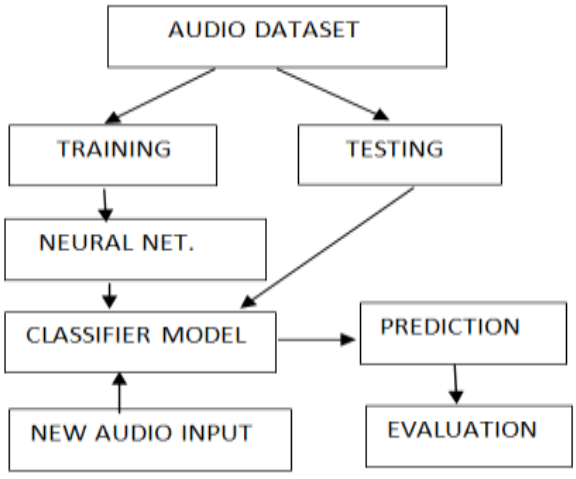

Fig -1: Block Diagram

3.1 WORKING

The audio dataset is in the form of .mp3 files. Every audio file has varied characteristics that can be exploited to segregate them into different classes or types. The process essentially involves converting the audio file into a spectral graph and analyzing the variations. The various features of the spectrum are spectral centroid, spectral bandwidth, spectral roll off.

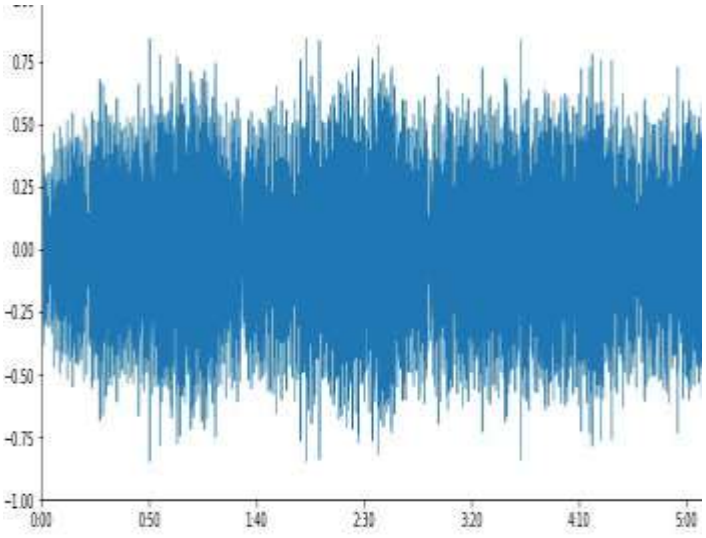

Fig -2: wave plot for an .mp3 file

The system utilizes the wave plot for individual audio files to extract these features and make use of it in the classification process. The ultimate output of the model is to predict which genre the audio will belong to for that process, a neural network is developed based on these features. Patterns in the features are the basis for classification of a music file into a genre. The system begins by a node network for relevant features like RMSE (Root mean square error), Chroma frequency coefficients, MFCCs (Mel frequency Cepstral coefficients) and spectral features to yield output as one of the 10 different genres viz blues classical country disco hip-hop jazz metal pop reggae and rock.

3.2 ACTIVATION FUNCTIONS

The model makes use of 2 activation functions namely RELU and SoftMax. The initial layers make use of RELU as the error rate can be high so the model may require to backtrack often. RELU is one of the widely used activation functions in neural networking models due to its backtracking abilities.

SoftMax is used in the output layer to map non normalized output of the previous layers to a probabilistic distribution over predicted output classes. In other words, it is used to determine the probability that an audio file belongs to a genre to classify it correctly.

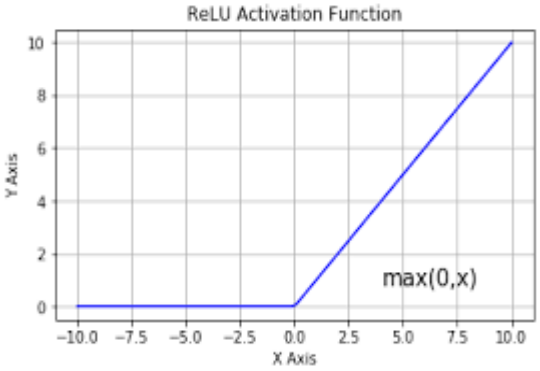

Fig -3: Graph for RELU function

The standard (unit) softmax function $\sigma : \mathbb{R}^K \to \mathbb{R}^K$ is defined by the formula

$$\sigma(\mathbf{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \text{ for } i = 1, \ldots, K \text{ and } \mathbf{z} = (z_1, \ldots, z_K) \in \mathbb{R}^K$$

Fig -4: Function definition for SoftMax

4. RESULTS

The results are compared by actually passing labelled input .mp3 files to the model. The evaluation process is based on the correct prediction the label genre of the audio file. 200 mp3 files were passed as input to the model in the testing phase. The trained model accepts input audio file, performs feature extraction and passes it through the network where a node is fired if the threshold expectancy is met. The testing phase yielded 2 types of results. 1.Correct prediction: When the predicted genre and the actual labelled genre matched.

2. Incorrect prediction/ Misclassification: The predicted genre is any genre out of the 10 except the one which is the actual label of the audio file.

The test loss was found to be minimal and the model performed decently with formidable accuracy.

```
model.compile(optimizer='adam',
              loss='sparse_categorical_crossentropy',
              metrics=['accuracy'])
```

```
Epoch 50/50
600/600 [==============================]
200/200 [==============================]
Test loss: 1.8840042781829833
Test accuracy: 0.68
```

Fig -5: Sample classification output for a .mp3 file

5. CONCLUSION

The model fulfills its primary objective of classification of an audio file into a specific genre. The system proves out to be exemplary for performing the classification without human intervention and thereby making it faster and easier process. From the results, it was observed that model could predict the genre of roughly 7 out of 10 audio files thereby rightly classifying them during the test phase. Misclassification mainly occurs due to wrong feature classification or complex audio tracks. The limitation of the model is that it can classify an audio file only in the provided 10 genres. An audio file that doesn't belong to any of these 10, is bound to be misclassified. The scope for improvement lies in extracting better features that are instrumental in determining the genre of an audio file more efficiently apart from those which are used in the model. The model in future can be combined with a sentiment analysis model to form a consolidated model to predict type of music the user would like to listen based on his mood.

REFERENCES:

[1] Automatic chord recognition for music classification and retrieval. In 2008 IEEE International Conference on Multimedia and Expo, ICME 2008 - Proceedings, 2008.

[2] Hareesh Bahuleyan. Music Genre Classification using Machine Learning Techniques. 2018