JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

MUSIC GENRE CLASSIFICATION USING NEURAL NETWORK

[1]Naman Kothari, [2]Pawan Kumar
[1]Student, [2]Assitant Professor
[1]Department of Information Technology
[1]Jain Knowledge Campus, Bangalore, Karnataka

Abstract: *Primarily aimed at creating an automated system for classifying music genre models. The first step was to find a good feature that clearly demarcates the boundaries of the genre. The most common characteristic that can be extracted from sound is the Mel Frequency Cepstral Coefficient (MFCC). MFCC representing the Mel Frequency cepstral coefficient. The input to CNN is a short-time Fourier transform of the audio signal. The output of the CNN is passed to another Deep Neural Network for classification. There are many ways to classify songs by genre using song libraries, machine learning techniques, input formats, and neural networks. The spectrogram generated by the time song server is used as a record for a neural network (NN). Develop a system to train a Convolutional Neural Network (CNN) using a deep learning approach based on the generated spectrogram. Preliminary experimental results using the GTZAN dataset when comparing two different network topologies show that the above two methods can effectively improve the classification accuracy, especially when compared with the second method.*

Index Terms - Classification of Music Genre, Convolutional Neural Networks, Feature Extraction, Mel Spectrogram, GITZAN dataset, Mel-Frequency Cepstral Coefficient.

I. INTRODUCTION

Music is categorized into subjective categories called genres. With the growth of the internet and multimedia systems applications that deal with the musical databases gained importance and demand for Music Information Retrieval (MIR) applications increased. Musical genres have no strict definitions and boundaries as they arise through a complex interaction between the public, marketing, historical, and cultural factors. This observation has led some researchers to suggest the definition of a new genre classification scheme purely for the purposes of music information retrieval Genre hierarchies, typically created manually by human experts, are currently one of the ways used to structure music content on the web. Automatic musical genre classification can potentially automate this process and provide an important component for a complete music information retrieval system for audio signals. We use audio data set which contains 1000 music pieces each of 30 seconds length. There are 100 pieces from each of the following genres: Blues, Classical, Country, Disco, Hip Hop, Jazz, Metal, Pop, Reggae, and Rock.

1.1 Machine Learning and Neural Networks:

Machine learning has been gaining popularity in recent years. Some of the various applications are more efficient than others. The most common types of learning algorithms are unassisted reading, supervised learning, supervised reading, and reinforcement reading.

Neural network (NN) is a way to learn the machine, and it is often effective when you remove key elements from a large data set and reflect these functions. First, the NN uses a training database to train the model. Once the model is trained, the NN can be applied to new or previously unselected data points to separate the data using the previously trained model.

A Convolutional Neural Network (CNN) is a kind of central network designed to process the same element as an image from multiple angles. CNNs can be used for dual classification and classification functions in several categories, the only difference being the number of output classes. For example, you can train an image breaker using an animal data set. A CNN is given a vector of pixel values in the image and contour segments defined by the vector (cat, dog, bird, etc.).

Deep Neural Networks (DNNs) are the most widely used diagnostic tools and help with Large Gene Expression Website (MFCC) training. Later released, these structures are used as training neurons. Separating music is considered a difficult task because of the selection and selection of acceptable sound elements.

1.2 Music Classification:

Music classification is a type of music retrieval function (MIR) where labels are assigned to music features such as genre, heart rate, and instruments. It is also associated with concepts such as musical similarities and musical tastes. People have known about music since the beginning of time. The concept of classification of music allows us to distinguish between different genres based on their composition and frequency of hearing. With the increasing variety of genres around the world, the classification of genres has recently become quite popular. Genre classification is an important step in building a useful reward system for this project. The ultimate goal is to create a machine learning model that can more accurately classify music samples into different genres. These audio files are classified according to the characteristics of the minimum frequency and time period.

II. RELATED WORK

In the studies of Music Genre Classification, GTZAN dataset is been used because it is an open-source dataset present for music with 10 different genre available to train and test the stereo channels had been then blended into one mono channel, and the track statistics became transformed right into a spectrogram the use of the SoX (Sound eXchange) command-line track software utility, which became then sliced into 128x128 pixel images, and the labelled spectrogram became used as inputs to the dataset, which became cut up into 70% education statistics, 20% validation statistics, and 10% check statistics. Following the primary 4 layers is a totally related layer wherein every output of the preceding layer is fed into every enter of the completely related layer. The CNN implementation seemed to be overfit because the accuracy for the education statistics became 97% as opposed to 47% for the check statistics. The study of the Spotify music dataset containing 228,159 songs in 26 genres and 18 features. To process data structures and perform data analysis, the Python programming language is used in conjunction with the Python Data Analysis Library (PANDAS). Cross-validate the Fold with free random conditions and compare 80% of the training data and 20% of the test data. In [6], they claim to have used convolutional neural networks (CNNs), which have been successfully used in deep learning, especially recently in computer vision and speech recognition. One specific example in the Music Information Retrieval (MIR) process is the classification of music genres. The Mel spectrogram as the input to the CNN to further reduce the dimension of the spectrogram. For Prediction and Training 1000s of music tracks (converted to Mel Spectrogram) are evenly divided into training, validation, and testing sets in a 5:2:3 ratio. During testing, all music (Mel Spectrogram) is divided into 3second segments with 50% overlap. The trained neural network then predicts the probability of each genre for each segment in [6]. Shazam describes the song's trademark as a large amplitude peak in the song's spectrogram. Deep learning, particularly the use of Convolutional Networks (CNNs), has recently been successfully used in computer vision and speech recognition. The GTZAN dataset is a popular music genre database for model submission. Thousands of songs in 10 genres including blues, classical, country, disco, hip-hop, jazz, metal, pop, reggae, rock, etc. All weights can be initialized against the input vector which will be a 128x128 pixel spectrogram.

III. METHODOLOGY

The main categories of music genres to achieve my goals in my research are datasets, data preprocessing, feature release of datasets, model training and validation.

Fig. 3.1: System Design Process

3.1. Dataset

The GTZAN database is used to enter data into the system, as it is a collection of free songs of various genres. The collection consists of thousands of audio tracks divided into 10 genres. This database includes blues, classical, country, disco, hip hop, jazz, metal, pop, reggae and rock. Music data from the GTZAN database was taken at 22050Hz and lasted about 30 seconds, for a total of 22020 x 30 = 661500 samples. For each smoothing window of 2048 samples with 1024 sample variance calculated during the study, all results below are evaluated for 10 runs using the cross-sectional accuracy selected as the performance metric.

Table 3.1: GTZAN Dataset Genre Distribution.

Genre	Number of Songs
Blues	100
Classical	100
Country	100
Folk	100
Hip-Hop	100
Jazz	100
Metal	100
Electronic	100
Reggae	100
Rock	100
TOTAL	**1000**

3.2. Data Pre-processing

In this step, each music signal is first converted from a waveform to a Mel spectrogram with a 23 MS time window and passed through the Librosa package. It then converts the chalk spectrogram to logarithm to set other chalk scale values in the same range. Because the Mel spectrogram is a biologically inspired image, it is easier to understand the PCA whitening method [6]. Training data and test data were distributed as 80% training data and 20% test data.

3.3. Feature Extraction from Dataset

The Python librosa module is used to extract feature vectors. This software is for audio analysis only. Extract each audio file and compute vector features. By roughly recording logarithmic spectral energy types on a choke frequency scale, MFCC incorporates the tonal function of a music signal. Choke Spectrogram: A choke spectrogram mimics the genetic makeup of humans by representing the temporal frequency of sound. Magnitude spectra are computerized from time series data and displayed on a chalk scale.

3.4. Training & Testing of Model

10,000 music tracks (converted to Mel Spectrograms) are equally divided into sets for training, validation and testing in a ratio of 5:2:3. The learning process is as follows.

- Select a subset of tracks in the training set.
- Select a 3 second continuous segment from all selected songs and pick a random starting point.
- The gradient is computed using a backpropagation algorithm with segments as input and source music labels as target genres.
- Update the weights with the gradient.
- Repeat the procedure until the classification accuracy of the cross-validation data set improves.

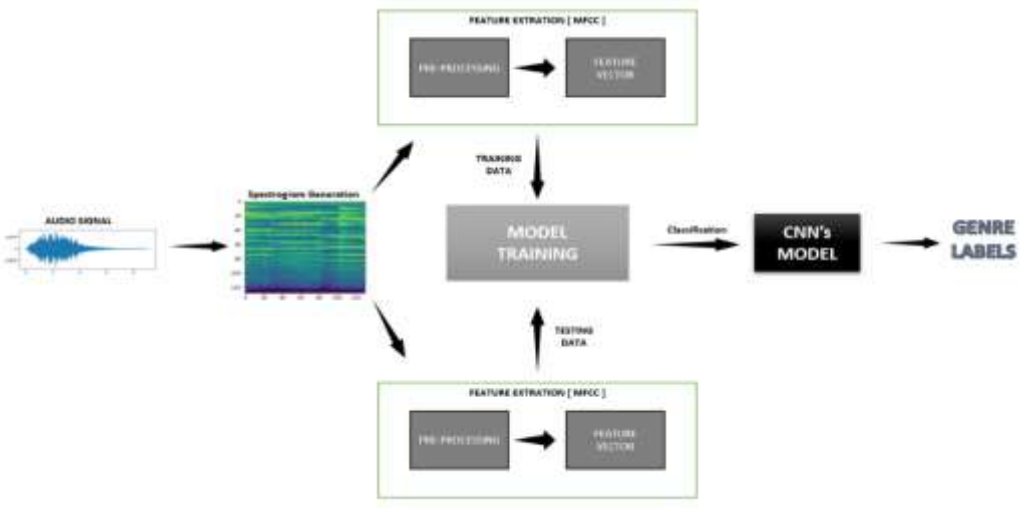

Fig.3.4.1: System Work flow Diagram

During testing, all music (mel-spectrogram) is split into 3 second pieces overlapping 50%. The trained neural network then predicts the probability of each genre for each section. The genre with the highest average probability is the one predicted for each piece of music.

3.5. Convolutional Neural Network Model

Convolutional Neural Networks (CNNs), like traditional multi-layer neural networks, consist of one or more dynamic layers and one or more fully connected layers. This step involves passing a matrix filter (eg 3x3) to the embedded image of size (image_width x image_height). The filter is first applied to the image matrix, then smart iterations of the elements between the filter and the image region are computed, and then a summary is computed to determine the element values. The model consists of four convolutional layers. Each layer consists of a convolution filter, a ReLU activation function, and a bulk integration layer for size reduction. Before we move on to the neural network, we have a planar layer and a fixed layer. A flat layer converts an image tensor to a vector. This vector is the input to the neural network. A stop layer is applied to avoid overload. A neural network consists of a dense layer of 512 nodes and an outgoing layer with nodes equal to the number of common classes.

IV. IMPLEMENTATION

Neural Networks has been fulfilled through several changes. The difference includes increasing the number of musicals in the genre and increasing the number of training spectrograms. It teaches the use of various music data sets as well as the pieces of genres used to teach NN. The operation of changing the activation function uses the various functions of the digital signal processing function when converting the MP3 file from the spectrogram and increases the number of CNN layers. Two music libraries have been used in this article. It consists of 1880 songs categorized in both genres. Converts the stereo channel to a mono channel and preparation for a musical data set using SoX (Sound eXchange) music utility application that converts musical data to spectrums. An example of a pop spectrogram is shown at the top. The next step in preparing the data set was to cut the large spectrogram into a 128-pixel wide PNG file. Each file represents 2.56 seconds of a given song. A representative example of a 128 × 128-pixel image spectrogram fragment is shown below. Both datasets consisted of 1880 different songs, each 3 minutes long. By dividing the song into spectrogram segments of 2.56 seconds, approximately 132,000 labeled spectrogram fragments were obtained. The labeled input dataset spectrogram was split into 70% training data, 20% validation data, and 10% test data. This is a variant. NN implemented a deep CNN using TensorFlow. Final CNN architecture. All weights were initialized using Xavier initialization and the input vector is a 128x128 pixel spectrogram. first six layers are a convolutional layer with a kernel size of 2x2 with a strike of 2 and a maximum pooling layer implemented after each successive layer. The first six layers are followed by a fully connected layer. Here, each output of the previous layer is fed to each input of the fully connected layer. This method produces an array of length 4096. Then apply a SoftMax layer to define 10 outputs, each representing a specific genre. The activation function used throughout the network is the Rectified Linear Unit (ReLU). The optimizer used the Adam optimizer. To limit overfitting, we implemented dropout with a probability of 0.5 during training.

V. RESULT AND DISCUSSION

5.1 Model Discussion

The suggested approach will be used to categories the music database into several genres. There are numerous databases that can be used to train the system. The GTZAN Database is used to enter data into the system because it is a collection of free accessible songs from many genres. The collection is made up of thousands of audio tracks divided into 10 different genres. This database includes blues, classical, country, disco, hip-hop, jazz, metal, pop, reggae, and rock music. Music data on the GTZAN database is taken at 22050 Hz and lasts for about 30 seconds, a total of 22020 x 30 = 661500 samples. For each a smooth window of 2048 samples, with a change of 1024 samples, as calculated during the study, all the results provided below are rated at more than ten runs, and the accuracy of the sections was selected as metric performance metrics. The librosa module in Python is used for Feature Vector Extraction. This software is used for audio analysis only. Each audio file is extracted and the vector feature is calculated. By recording approximately, the type of log energy spectrum on the Mel-frequency scale, MFCCs incorporate timbral features of the music signal. 10,000 music tracks (converted to mel-spectrogram) are evenly divided into training, validation, and testing sets in a 5:2:3 ratio.

5.2 Results for Convolutional Neural Network Model

Table 5.2.1: Experiments Result for Convolutional Neural Network

Training Accuracy	**98.7247**
Testing Accuracy	**88.4848**

VI. CONCLUSION

This article presents an application for classifying music genres using neural network methods. There are many ways to extract audio features and we decided that MFCC was the best for this project. I used KNN and CNN algorithms to train the model and perform classification on the data set. This article shows a music genre classification system based on neural networks. Music classification is a type of music information retrieval (MIR) activity that labels musical elements such as genre, mood, and instrument. The Python-based librosa library helps extract features, providing parameters suitable for network training. Consequently, this system looks promising for classifying large music databases into related genres.

The CNN module was used as a feature extractor for learning intermediate and advanced features from spectrograms. CNN is a biologically inspired multi-layered variant of the perceptron. In this work, I work with audio sets, which are large databases of human annotated sounds. A collection of music divided into several genres were used. Then crop a large 128 wide PNG spectrogram representing 2.56 seconds of this song. The GTZAN data set is used here. There are 10 classes (10 music genres) with 1000 audio tracks each. All tracks in .wav format. It includes audio tracks from 10 genres listed below. Some genres are blues, classical, country, disco, hip hop, jazz, metal, pop, reggae and rock. The GTZAN dataset has long been used as a standard for defining music genre classifications, and the MFCC spectrogram is also used for track preprocessing.

REFERENCES

[1] Nirmal M R, "Music Genre Classification using Spectrograms", *International Conference on Power, Instrumentation, Control and Computing (PICC) – 2020.*

[2] Naman Kothari, "Literature Survey on Music Genre Classification Using Neural Networks", IRJET, Volume-09, Isseue 02, Feb-2022.

[3] De Rosal Ignatius Moses Setiadi, Dewangga Satriya Rahardwika, Candra Irawan, Desi Purwanti Kusumaningrum, "Comparison of SVM, KNN, and NB Classifier for Genre Music Classification based on Metadata", *International Seminar on Application for Technology of Information and Communication (iSemantic) – 2020.*

[4] Nikki Pelchat and Craig M. Gelowitz, "Neural Network Music Genre Classification", *IEEE Canadian Conference of Electrical and Computer Engineering (CCECE) – 2019.*

[5] K. Meenakshi, "Automatic Music Genre Classification using Convolution Neural Network", *International Conference on Computer Communication and Informatics (ICCCI)– 2018.*

[6] Mingwen Dong, "Convolutional Neural Network Achieves Human-level Accuracy in Music Genre Classification", *arXiv:1802.09697v1 [cs.SD] – 2018.*

[7] Hareesh Bahuleyan, "Music Genre Classification using Machine Learning Techniques", arXiv:1804.01149v1 [cs.SD] – 2018.

[8] Weibin Zhang, Wenkang Lei, Xiangmin Xu, Xiaofeng Xing, "Improved Music Genre Classification with Convolutional Neural Networks", *Interspeech – 2016.*

[9] Balachandra K, Neha Kumari, Tushar Shukla, Kumar Satyam, "Music Genre Classification for Indian Music Genre", *IJRASET* – 2021.

[10] Nicolas Scaringella, Giorgio Zoia, Daniel Miynek, "Automatic Genre Classification of Music Content", *IEEE Signal Processing Magazine* – March 2006.

[11] Deepanway Ghosal, Maheshkumar H. Kolekar, "Music Genre Recognition using Deep Neural Network and Transfer Learning", in *Interspeech* vol. *28, no. 24, pp. 2087-2097, September 2018.*

[12] Jose J. Valero-Mas, Antonio Pertusa," End-to-End Optical Music Recognition Using Neural Networks", 18th *International Society for Music Information Retrieval Conference – 2017.*

[13] Prasenjeet Fulzele1, Rajat Singh2, Naman Kaushik3," A Hybrid Model for Music Genre Classification Using LSTM & SVM", *ICCC – 2018.*

[14] George Tzanetakis, Perry Cook," Musical Genre Classification of Audio Signals", *IEEE Transactions on Speech & Audio Processing, vol. 10, No. 5, July 2002.*