International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 03 | Mar 2021 www.irjet.net p-ISSN: 2395-0072

A Survey on Music Genre Classification using Machine Learning

Prof. Shweta Koparde[1], Vaishnavi R Bhadgaonkar[2], Kalyani N Patil[3], Gauri N Basutkar[4], Dhanashri D Gayke[5]

[1]Professor, Dept. of Computer Engineering, Pimpri Chinchwad College of Engineering and Research, Ravet
[2,3,4,5]Student, Dept. of Computer Engineering, Pimpri Chinchwad College of Engineering and Research, Ravet

---

Abstract - *Music has become the most favorable area nowadays especially in youth. Most of the people tend to listen music of certain genre such as classical, hip-hop or disco and want a user-friendly way to classify the music as per their preferences. Due to this "Music Genre Classification" came into picture. Music genre classification is a complex task in music information retrieval (MIR) due to selection and extraction of suitable features. The Machine learning models have been shown to be capable of solving these kinds of real life problems. Music genre classification can be implemented using various machine learning algorithms. In the proposed system, we are using a deep learning technique .i.e. convolution neural network (CNN) for classifying the music in various genres. CNNs are used to solve image pattern recognition tasks. While analyzing music, acoustic feature extraction is the most crucial task. In proposed system, model is trained over GTZAN dataset.*

Key Words: **Deep Learning, Convolution Neural Network, classification, neural network.**

1. INTRODUCTION

Now-a-days, variety of songs are available and in various genres. People find these songs soothing, encouraging and cheerful. Studies show that soothing music triggers relaxation and improve well-being of our mind, helps reduce stress and anxiety. Also, due to globalization, there are significant changes in the Music Industry because various people like musicians, music producers have taken influence of various music art forms present around the world to produce soulful music. So varieties of songs are available to the users to choose from.

There are number of music streaming providers like Spotify, Gaana, Prime Music, Youtube Music, etc. which with the help of new emerging technologies have improved their songs recommendations and segregation system. These providers use Machine Learning to improve and enhance user experiences and have made streaming music really easy for the customers.

Machine Learning in short can be defined as the name suggests is training of a machine i.e. a computer algorithm. We make this machine capable of learning various things and without doing any kind of explicit programming. It's an interesting branch of Artificial Intelligence where systems i.e. machines learn through various data available called as datasets, identify various patterns and take decisions with very minimal human interaction. Various algorithms are present with their advantages and disadvantages suiting a particular real time issue and with the help of them the system learns, classifies and predicts the decision. There are three main categories in which Machine Learning can be classified:

a. Supervised Learning :

Here, the system i.e. the algorithm learns from examples and various responses like string labels or numerical values, which can be used later to predict the output when given a new situation for prediction is called as Supervised Learning. It develops predictive modal using both the input as well as the output. Classification and Regression falls under supervised learning.

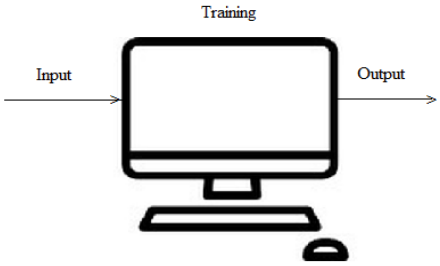

Fig 1 – Supervised Learning

b. Unsupervised Learning :

In Unsupervised learning, the algorithm learns without any responses like labels, etc. just plain examples are there to help the algorithm identify the data patterns and make decisions. It just groups and interprets data based on the input provided to the algorithm. Clustering falls under unsupervised learning.

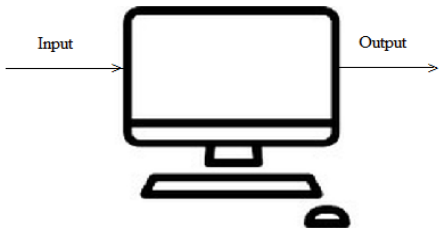

Fig 2 – Unsupervised Learning

c. Reinforcement Learning :

In Reinforcement learning, no labels are used same as unsupervised learning. However, a concept called reward i.e. feedback is used. This feedback can be positive or else negative and is given with a example for the system to come to a decision. It's similar to a human learning from trial and error.

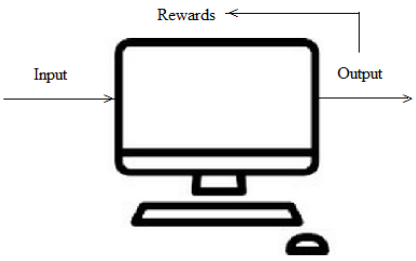

Fig 3 – Reinforcement Learning

Also, a huge variety of songs are available in the Market and there is a need of classification of music according its genres. Let's consider an example, if we are having an event at our house, we will always give the preference to the songs matching to that event like if there is a party, rock songs or jazz songs are always the choice. So due to this, it is important and very useful in content based music retrieval and music distribution. Also, due to this genre classification we can understand which are the favourite genres of users and depending on those we can give recommendations to the user.

When a user downloads a song, it gets saved into the downloads folder and if he goes on downloading multiple songs, all the songs will be cluttered there and the user has to segregate them manually into various folders genre-wise which is tedious and time consuming process. Hence we are proposing a system to automate this for users to ease the process of music genre classification with improved performance and accuracy using a machine learning approach.

When a user downloads a song into his system, he/she has to manually save the song into folder of its appropriate genres. This process is tedious and time consuming. Due to this we are proposing our system, where the user just has to upload the song on the website and the song will get segregated into its respective folder

2. LITERATURE SURVEY

S. Vishnupriya and K. Meenakshi[10] in their paper "Automatic Music Genre Classification using Convolutional Neural Network" published in year 2018 proposed an automatic music genre classification system using Convolution Neural Network(CNN). They used GTZAN dataset as input to their system. Feature extraction is an important task which is performed in pre-processing and they used Mel Frequency Cepstral Coefficient (MFCC) as feature vector. Librosa package is used for feature vector extraction in python. The size of feature vector is 599x13x5 for MFCC. From this database, input is then given to generate spectrograms in the training phase.The learning accuracy obtained using MFCC in this paper is 76%.

Yandre M.G. Costaa, Luiz S. Oliveira b, Carlos N. Silla Jr. c[3] in their paper "An evaluation of Convolutional Neural Networks for music classification using spectrograms" published in year 2017 proposed music genre classification system by using Convolution Neural Network and Support Vector Machine algorithm. They used three dataset: Latin Music Dataset [LMD], ISMIR, African Music Database. LMD consisted of three 3,277 full-length music pieces of 10 different genres. ISMIR 2004 consisted of 1,458 music pieces. This dataset was divided into 50% training and 50% testing.In this paper output of both algorithms is combined to get more accuracy. The result obtained is the fusion of CNN and SVM which gives 83% Accuracy.

Nikki Pelchat and Craig M. Gelowitz[2] in their paper "Neural Network Music Genre Classification" published in year 2019 proposed a music genre classification system using a machine learning approach that is Convolutional Neural Network (CNN) algorithm. They used a dataset consisting of 1880 songs of different genres. Length of each song in dataset was of 3 minutes. They generated 132,000 labelled spectrogram images by dividing song into 2.56 second segment spectrograms .Then, they splitted dataset into 70% training data, 20% validation data, and 10% test data. These spectrogram images are given as input to CNN algorithm which then classifies the song according to their genre. The accuracy obtained in this paper is 67%.

Elizabeth Nurmiyati Tamatjita and Aditya Wikan Mahastama[4] published a paper .i.e "Comparison of Music Genre Classification Using Nearest Centroid Classifier and k-Nearest Neighbours" in 2016. In this paper they compared music genre classification using Nearest Centroid Classifier (NCC) and K-Nearest Neighbours (KNN). Here, Zero Crossing Rate (ZCR), Silent Ratio (SR) and Average Energy (E) features are extracted from audio files in order to classify songs genre. Their dataset included songs of 12 different genres. Selected features .i.e. ZCR, SR, E are giving the classification with maximum accuracy. The accuracy obtained in this paper is 56.3% with k-Nearest Neighbours (k-NN).

3. DATASETS

3.1 ISMIR 2004

ISMIR 2004[11] dataset is a collection of audio songs used for Genre Identification which was created for a contest organized by the Music Technology Group (Universitat Pompeu Fabra). Songs were from 8 different genres namely classical, electronic, jazz & blues, metal, punk, rock, pop and world and from them 6 classes are formed which are electronic, Jazz/Blues, world, classical, metal/punk and rock/pop. This dataset consist of 320 classical, 115 electronic, 26 Jazz/Blues, 45 Metal/punk, 101 rock/pop and 122 world songs. All the songs are of MP3 format and it is divided into 3 folders which are Training, Development and Evaluation. Each folder has 6 classes and 729 files.

3.2 Latin Music Dataset

Latin Music Dataset (LMD)[12] includes 3,227 music songs of different 501 artists. Songs are of 10 different genres namely Bachata, Bolera, Axe, Merengue, salsa, Gaucha, pagode, Forro, Tango and Taneja. Dataset includes songs with significant similarity with respect to rhythmic structure, instrumentation and harmonic content.

3.3 African music dataset

African music dataset[13] is of royal museum of central-Africa (RMCA) in Belgium. Music songs belong to 4 different categories country, function, ethnic group and instrumentation. Characteristics of this music collection are quite different as compared to typical western music dataset.

3.4 GTZAN

GTZAN[6] i.e. Genre Collection dataset from MARSYAS (Music Analysis Retrieval and Synthesis for Audio Signals) contains 1000 songs each of 30 seconds long and has 10 genres like hip-hop, rock, classical, blues, country, disco, jazz, reggae, pop, etc. This is the most used dataset available publicly for evaluation in research for music genre recognition (MGR) for Machine Learning.

4. ARCHITECTURE

4.1 Proposed System

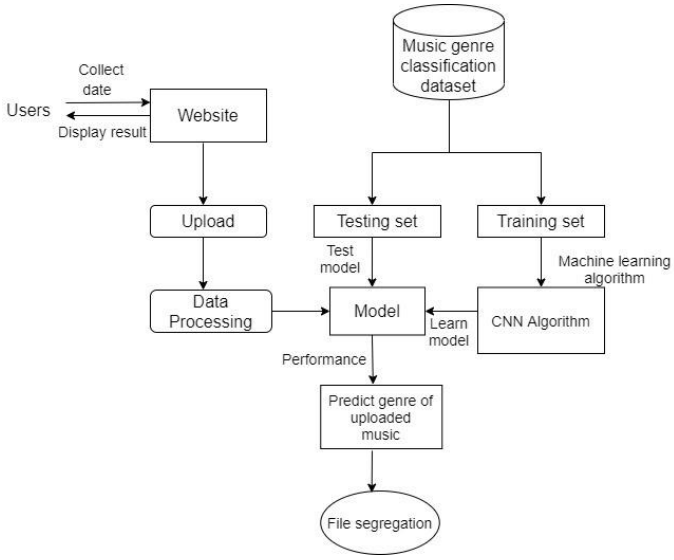

Fig 4 – Architecture Diagram

In the proposed model, we are developing a Music Genre Classification System which will classify the songs as per their genres and we are using the GTZAN dataset. After reading the dataset, mel spectrograms are created which is a feature extraction of our song. Pixel values of spectrograms are stored in a file and this file is passed as an input to the CNN algorithm along with the song to be classified. With the help of pixel values CNN algorithm classifies the song into their respected genre.

Also, we are proposing the next module in which our system will automatically classify/put the output of algorithm i.e. classified songs into their respective folder as per their genre; this will help to get a list of songs according to genre. This classification of songs into separate folders according to genre will help user to choose any genre according to the mood or an event. Also, managing of the songs properly genre-wise will help user in efficiently handling the music and in choosing the list of songs.

Therefore, this application has following main modules:

1. Spectrogram File Generation
2. Music Genre Classification
3. User interface i.e. Website
4. Store the song in appropriate genre folder

4. 3 Feature Extraction

Here, Mel Scale plays an important part. It relates perceived frequency, or pitch, of a pure tone to its actual measured frequency. A frequency which is measured in Hertz (f) can be converted to Mel Scale using the below given formula[19]:

$$\mathrm{Mel}(f) = 2595 \log\left(1 + \frac{f}{700}\right)$$

Feature extraction in music genre classification can be done by following ways –

a. Spectrograms -

Spectrogram is one of the ways to visually represent the strength or loudness of a signal as it varies with time at various frequencies present in a particular waveform. Spectrograms are generally 2-Dimensional graphs having third variable which is represented by colour. Optical spectrometer is used to generate spectrogram. Horizontal axis represents time and Vertical axis represents frequency or pitch or tone. The third dimension represents audio's energy or amplitude or loudness.

To construct the spectrogram of audio, Signal Analyser follows these steps[18]:

i) Audio is divided into equal-length segments. The segments should be short enough that the frequency content of the audio doesn't change significantly within a segment. The segments may or may not overlap.

ii) Then each segment is windowed and its spectrum is computed to get the Short-time Fourier Transform.

iii) Display spectrogram segment by segment, the power of each spectrum in decibels.

b. MFCC –

Mel Frequency Cepstral Coefficients (MFCCs)[17] are mainly used for feature extraction. It is used in automatic speech recognition. Prior to MFCCs, Liner Prediction Coefficients (LPCs) and Linear Prediction Cepstral Coefficient (LPCCs). First step in automatic speech recognition system is to extract all the features .i.e. identify components that are good for identifying linguistic content and removing all other stuffs like background noise, emotion, etc. The main point is that it identifies shape of the vocal tract generated by human and this shape determines what sounds are generated. If this shape is determined accurately, we have a accurate representation of sound being produced. So the job of MFCCs is to accurately represent all this information.

Following are the steps in creating a MFCC:

i) Frame the signal into short frames of 20-40 ms. 25ms is considered as a standard.

ii) For individual frame, calculating periodogram estimate of the power spectrum - Here periodogram estimate of power spectrum is taking Discrete Fourier Transform of the individual frame.

iii) Apply Mel-spaced filterbank - This is a set of 20-40 (26 being the standard) triangular filters which we apply to step 2's output which is a periodogram power spectral estimate. Here, filterbanks have vector length of 257 and has 26 vectors in total.

iv) Take logarithm of filterbank - Here we are taking log of each 26 energies from step 3.

v) Final step is to find discrete cosine transform of log filterbank energies keeping DCT coefficients 2-13 and discarding the rest.

c. Mel-Spectrograms –

A Mel-Spectrogram is a type of spectrogram in which frequencies are converted into mel scale. Here, x-axis represents time and y-axis represents mel scale. The waveforms of audio files are passed through mel filter banks which then gives mel spectrogram. Librosa Library in Python can be used to generate Mel-Spectrograms.

5. APPLICATIONS

Mall:

Music is played continuously in the malls, and selection of right music to be played is hectic as well as time consuming work. So here, our system helps to choose the song according to any occasion or event.

Restaurant:

In a restaurant, choosing the right music is an important task when it comes to various occasions as per customer's demand; our system will help to choose a particular genre song for the same.

Airport:

Music is played in the airports for the entertainment of people as they wait for hours due to various reasons, so our system will help to choose the song as per the requirements.

6. CONCLUSIONS

In this system, we are proposing an efficient Music Genre Classification System which can classify the uploaded audio file into various genres. We are using GTZAN dataset with 1000 songs so as to train our system. Features are extracted using Librosa module provided in Python and after this spectrograms are generated. We are using CNN Algorithm for the classification purpose. The output of CNN classifier i.e. classified song is then segregated into their respective folder.

While concluding we can say that when we compared CNN to various algorithms like K Nearest Neighbor, FFNN, etc. CNN Algorithm gave better accuracy in classifying the genre. Also, we are using Mel Spectrums instead of MFCC (which is less accurate) for converting the audio signals into .png format images which are given as the input to our CNN model.

ACKNOWLEDGEMENT

Here, we express our sincere thanks to our project guide Prof. Shweta Koparde and the Head of the Department Prof.

Dr. Archana Chaugule for their constant support and for providing guidance. We are also thankful to all other researchers for their publications.

REFERENCES

[1] S. Vishnupriya, K. Meenakshi, "Automatic Music Genre Classification using Convolution Neural Network", 2018 International Conference on Computer Communication and Informatics (ICCCI), January 2018.
[2] Nikki Pelchat, Craig M. Gelowitz, "Neural Network Music Genre Classification", Canadian Journal of Electrical and Computer Engineering, Volume: 43, Issue: 3, Summer 2020
[3] Yandre M.G.Costa, Luiz S.Oliveira, Carlos N.Silla Jr., "An evaluation of Convolutional Neural Networks for music classification using spectrograms", Applied Soft Computing, Volume 52, March 2017, Pages 28-38
[4] Elizabeth Nurmiyati Tamatjita, Aditya Wikan Mahastama, "Comparison of Music Genre Classification Using Nearest Centroid Classifier and k-Nearest Neighbours", 2016 International Conference on Information Management and Technology (ICIMTech), 18 May 2017
[5] Machine Learning GeeksforGeek - https://www.geeksforgeeks.org/machine-learning/
[6] GTZAN Dataset Site : http://marsyas.info/downloads/datasets.html
[7] Musical Genre Classification with Convolutional Neural Networks by Leland Roberts : https://towardsdatascience.com/musical-genre-classification-with-convolutionalneural-networks-ff04f9601a74
[8] Understanding the Mel Spectrogram by Leland Roberts - https://medium.com/analyticsvidhya/ understanding-the-mel-spectrogram-fca2afa2ce53
[9] Convolutional Neural Network Tutorial by Simplilearn - https://www.simplilearn.com/tutorials/deep-learningtutorial/convolutional-neural-network
[10] Getting started with Django - https://www.djangoproject.com/start/
[11] ISMIR2004 Genre dataset - https://www.upf.edu/web/mtg/ismir2004-genre
[12] The Latin Music Database (LMD) - https://sites.google.com/site/carlossillajr/resources/the-latin-music-database-lmd
[13] Music of Africa dataset - https://research.google.com/audioset/dataset/music_of_africa.html
[14] Understanding Audio data, Fourier Transform, FFT and Spectrogram features for a Speech Recognition System by Kartik Chaudhary - https://towardsdatascience.com/understanding-audio-data-fourier-transform-fft-spectrogram-and-speech-recognition-a4072d228520
[15] Music Genre Classification Report by Derek A. Huang, Arianna A. Serafini and Eli J. Pugh from Stanford University - http://cs229.stanford.edu/proj2018/report/21.pdf
[16] François Chollet et al. Keras - https://keras.io, 2015.
[17] Mel Frequency Cepstral Coefficient (MFCC) tutorial- http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/
[18] Spectrogram computation in Signal Analyzer - https://www.mathworks.com/help/signal/ug/spectrogram-computation-in-signal-analyzer.html
[19] The dummy's guide to MFCC - https://medium.com/prathena/the-dummys-guide-to-mfcc-aceab2450fd