Automatic Music Genre Classification using Convolution Neural Network

VISHNUPRIYA S
BIGDATA ANALYTICS
SRM UNIVERSITY
Chennai,india
vishnupriya_sabareenathanp@srmuniv.edu.in

K.MEENAKSHI
INFORMATION TECHNOLOGY
SRM UNIVERSITY
Chennai,india
meenakshi.k@ktr.srmuniv.ac.in

Abstract— **Music Genre classification is very important in today's world due to rapid growth in music tracks, both online and offline. In order to have better access to these we need to index them accordingly. Automatic music genre classification is important to obtain music from a large collection. Most of the current music genre classification techniques uses machine learning techniques. In this paper, we present a music dataset which includes ten different genres. A Deep Learning approach is used in order to train and classify the system. Here convolution neural network is used for training and classification. Feature Extraction is the most crucial task for audio analysis. Mel Frequency Cepstral Coefficient (MFCC) is used as a feature vector for sound sample. The proposed system classifies music into various genres by extracting the feature vector. Our results show that the accuracy level of our system is around 76% and it will greatly improve and facilitate automatic classification of music genres**.

Keywords— Music Genre Classification, Deep Learning, Convolution Neural Network, classification, neural network.

I. INTRODUCTION

Downloading and purchasing music from online music collections has become a part of the daily life of probably a large number of people in the world. The users often formulate their preferences in terms of genre, such as hip hop or pop or disco. However, most of the tracks now available are not automatically classified to a genre. Given a huge size of existing collections, automatic genre classification is important for organization, search, retrieval, and recommendation of music.

Music classification is considered as a very challenging task due to selection and extraction of appropriate audio features. While unlabeled data is readily available music tracks with appropriate genre tags is very less. Music genre classification is composed of two basic steps: feature extraction and classification. In the first stage, various features are extracted from the waveform. In the second stage, a classifier is built using the features extracted from the training data. There has been many approaches that are used for the classification of music into different genre. With huge amount of music available in the internet it is needed for an automatic music genre classification. Each implementation uses various types of feature extraction. Some can take the timber, rhythm etc. as the classifying parameter, while some others take pitch, timber, beat etc. The types of features extracted varies from person to person. The Deep Neural Network (DNN) is a most widely used in classification problems and it is helpful in training huge database.

We propose a novel approach for the automatic music genre classification using Convolution Neural Networks (CNN). The features from the music are extracted. They are called as Mel Frequency Cepstral Coefficients (MFCC) for each song. They are obtained by taking the Fourier transforms of the signal, then taking the logarithmic of the power values and then taking the cosine transforms. The detailed explanation will be done in the forthcoming sessions. These extracted features then acts as the inputs to the neuros for training. For our work, we analyze music from ten various genres. The whole implementation is done in python programming language. The average accuracy obtained by using the MFCC feature vectors is 76%.

The rest of the paper is organized as follows. The literature review is presented in Section 2. The proposed system is presented in Sections 3, including feature vector extraction methodology, introduction to CNN and training and classification. Experimental results are described in detail in Sections 4. Finally, conclusion and future work are given in Sections 5.

II. LITERATURE REVIEW

Many researchers have worked on studying the musical parameters and the methods to classify them into different genres. This section will primarily look into some of the research already carried out. Tzanetakis and Cook [1] pioneered their work on music genre classification using machine learning algorithm. They created the GTZAN dataset and is to date considered as a standard for genre classification. Changsheng Xu et al. in [2] have shown how to use support vector machines (SVM) for this task. Authors used supervised learning approaches for music genre classification. Scaringella et al. [3] gives a comprehensive survey of both features and classification techniques used in the music genre

978-1-5386-2238-4/18/$31.00 ©2018 IEEE

classification. Riedmiller et al. in [4] used unsupervised learning creating a dictionary of features. Scheirer in [5] describes a real-time beat tracking system for audio signals with music. In this model, a filter bank is coupled with a network of combination filters that trace the signal periodicities to yield an impression of the main beat and its potency. A real-time beat tracking system based on a multitudinous agent architecture that trace several beat hypotheses in parallel is described in [6]. Tao in [7] shows the use of restricted Boltzmann machines and arrives to better results than a generic multilayer neural network by generating more data out of the initial dataset, GTZAN. In this paper a data distribution problem in the dataset is explained and it shows that it makes it hard to accurately classify more than 4 classes using only the GTZAN dataset. For song preprocessing, this paper suggests the use of MFCC spectograms as well. Aaron et al. [8] use MFCC spectograms to preprocess the songs. This work doesn't focus on genre recognition, but on song similarly for music recommendation. However, it was worth mentioning for their use of convolutional neural networks with ReLU activation on song clips preprocessed as MFCC spectrograms. Gwardys et al. [9] show an interesting approach involving transfer learning. They initially train the model on ILSVRC-2012 [10] for image recognition and then reuse 1 the model for genre recognition on MFCC spectograms. The architecture used in this article consists of five convolutional layers, the first two and the last one with max pooling as well. In the end, three fully connected layers. Authors in [11] used midi, pitch and duration as the features of the music to perform the classification to obtain good results.

With these following literatures in mind we propose a system for the automatic classification of music into different genres. This is explained in the following section. Proposed system design

III. PROPOSED SYSTEM

A. Music genre classification system

The proposed system is used for classifying the music database into different genres. There are many database available for training the system. The literature shows GTZAN and Million Song Dataset (MSD) are most widely used. Here in our approach we plan to use the MSD. This is freely- available collection of songs based on different genres. The dataset consists of audio tracks totally of 280 GB. This really is a tedious process to extract features from these huge set of music.

Table 1. Dataset genre distribution

Genre Number of Records	Genre Number of Records
blues	100
classical	100
country	100
disco	100
Hip-hop	100
jazz	100
metal	100
pop	100
reggae	100
rock	100
TOTAL	1000

B. Proposed System Design

The steps followed in our methodology is shown in Fig 1. Initially the database of the music is created. Then each song has to go through a preprocessing stage.

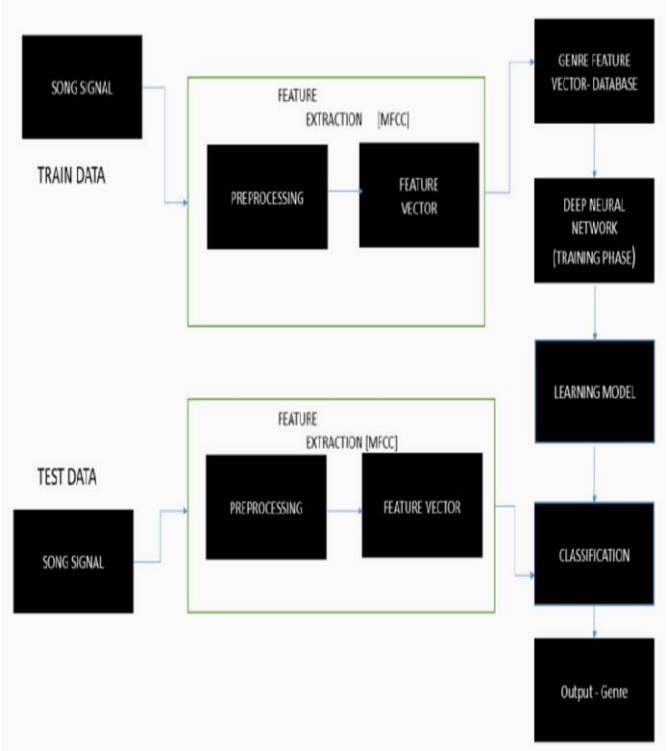
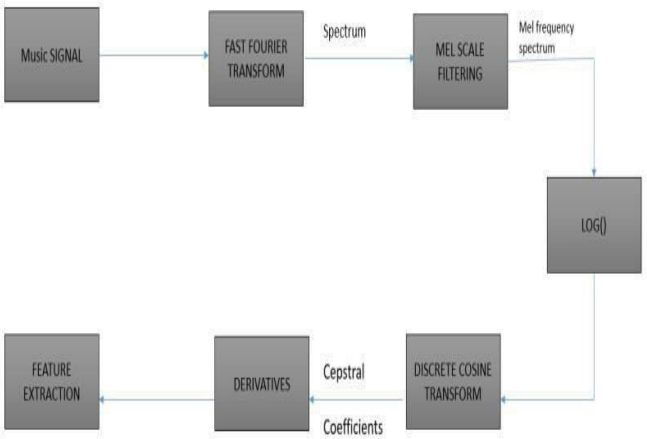


Fig. 1. Proposed system generalized block diagram

Feature Vector Extraction is done using the librosa package inpython. This package specifically used for the audio analysis. Each audio file is taken and from that the feature vector is extracted. The extracted feature vector is called as MFCC. The MFCCs encode the timbral properties of the music signal by encoding the rough shape of the log-power spectrum on the Mel-frequency scale. MFCC is calculated using the following steps as shown in Fig 2.



2.

Fig. 2. Feature Vector Creation steps

The music signal is applied with Fourier Transforms which converts the signal to the frequency spectrum. Then we do the Mel Scale Filtering for obtaining the Mel Frequency Spectrum. Then then is passed to a stage where the Log of the power is taken. The following is then passed on to block which takes the cosine transforms of the following signal and the feature vectors are obtained. Here we obtain two types of feature vectors, one is Mel Spectrum with 128 coefficients and another is MFCC with 13 coefficients. The comparison is done in the following section.

C. Convolutional Neural Network

A Convolutional Neural Network (CNN) it consist of one or more convolutional layers and then proceeded by one or more fully connected layers as in a standard multilayer neural network. Each neuron receives inputs from the feature vectors and then they are dot product with the weights which are then passed on to the next layers and optionally follows it with a non-linearity.

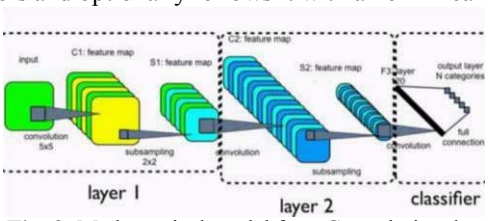


Fig. 3. Mathematical model for a Convolutional neural network

We use three main types of layers to build ConvNet architectures: Convolutional Layer**,** Pooling Layer, and Fully Connected Layer. Another important concept of CNNs is maxpooling, which is a form of non-linear down-sampling. Maxpooling partitions the input signals into a set of nonoverlapping matrix and, for each such sub-region, outputs the maximum value.

IV. EXPERIMENTAL SETUP AND RESULT

In the proposed model we use a supervised learning approach. The songs from the various genres go through the pre-processing phase where the Mel Frequency spectrum is taken as well as the MFCC is also calculated. These feature vectors are then stored into the database. The database thus obtained is the MFCC with 10 array size for genre. The input is 1000 songs, with 10 labels. The feature vector size is 599x128x 2 for Mel Spec and 599x13x5 for MFCC. The label size is 1000x10. Every song corresponds to a particular label. Before feeding the data to the neural network the data needs to be shuffled as we need a good form of generalization. For training 800 song features are taken and the remaining 200 is taken for testing. The training parameters are shown in table

Table 2: Network parameters for training and classification

Parameters	Values
learning_rate	0.001
training_iters	1,00,000
batch_size	64
display_step	1
train_size	800
n_input – MEL SPEC	599 * 128*2
n_input - MFCC	599 * 13 * 5
n_classes	10
dropout	0.75

The accuracy of the model is calculated using

$$\textbf{ACCURACY} = \frac{\textbf{No of songs correctly classified}}{\textbf{Total no of songs}} * 100$$

The evaluation was done using the Anaconda package for python. Tensorflow package was used for deep-learning. The system configuration on which the algorithm was implemented was Intel Xeon CPU E5-2630 v4 with 2.20GHz, 10 Core(s) and 20 Logical Processors, 32GB of RAM. The number of iterations was increased in steps of 10,000 starting from 10,000 to 1, 00,000. The learning accuracy tested with Mel Spec feature vector and MFCC feature vector was found to be 76% and 47 % respectively shown in Fig 5. MFCC takes less time for converging whereas Mel Spec is more time consuming for learning. Once the model is created the prediction takes less time.

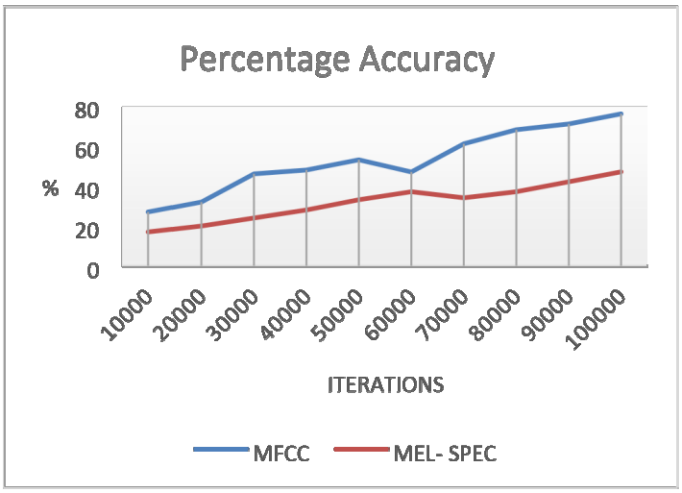


Fig. 5. Percentage of Accuracy

V. Conclusion and Future Work

This work shows provides a Convolution Neural Network based automatic music genre classification system. The feature vectors are calculated using Mel Spectrum and MLCC. The python based librosa package helps in extracting the features and thus helps in providing good parameters for the network training. The learning accuracies are shown to be 76% and 47% for Mel Spec and MFCC feature vectors respectively. Thus this methodology is promising for classification of huge database of songs into the respective genre.

The future work will focus on developing the system further to classify the songs based on mood. This will be helpful in finding out which kind of music can reduce stress in a person while listening to it. This be helpful in music therapy which can be used for playing a particular music depending on the person's stress level. This work needs to be further extended for such a system.

References

[1.] G. Tzanetakis and P. Cook. Musical genre classification of audio signals. Speech and Audio Processing, IEEE Transactions on, 10(5):293–302, Jul 2002.

[2] Chandsheng Xu, Mc Maddage, Xi Shao, Fang Cao, and Qi Tan, ―Musical genre classification using support vector machines‖, IEEE Proceedings of International Conference of Acoustics, Speech, and Signal Processing, Vol. 5, pp. V-429-32, 2003.

[3] N. Scaringella, G. Zoia, and D. Mlynek, ―Automatic genre classification of music content: a survey‖, IEEE Signal Processing Magazine, Vol. 23, Issue 2, pp. 133–141, 2006.

[4] Jan Wülfing and Martin Riedmiller, ―Unsupervised learning of local features for music classification‖ ISMIR, pp. 139–144, 2012

[5] E. Scheirer, "Tempo and beat analysis of acoustic musical signals," J. Acoust. Soc. Amer., vol. 103, no. 1, p. 588, 601, Jan. 1998.

[6] M. Goto and Y. Muraoka, "Music understanding at the beat level: Real-time beat tracking of audio signals," in Computational Auditory Scene Analysis, D. Rosenthal and H. Okuno, Eds. Mahwah, NJ: Lawrence Erlbaum, 1998, pp. 157– 176.

[7] T. Feng. Deep learning for music genre classification. 2014. [8] B. S. Aaron van den Oord, Sander Dieleman. Deep contentbased music recommendation. 2013.

[9] D. G. Grzegorz Gwardys. Deep image features in music information retrieval. 2014 10.O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A.

[10] Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision (IJCV), 115(3):211–252, 2015.

[11].Eve Zheng, Melody Moh, Teng-Sheng Moh, Music Genre Classification: A N-gram based Musicological Approach. 7th International Advance Computing Conference, 672-677, 2017.

[12].http://marsyas.info/downloads/datasets.html

[13].https://labrosa.ee.columbia.edu/millionsong/