iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 Issue: VI Month of publication: June 2021

DOI: https://doi.org/10.22214/ijraset.2021.35965

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VI Jun 2021- Available at www.ijraset.com

Music Genre Classification Using Deep Learning Techniques

M.K.N.Haq[1], B. Venkata Sai[2], J. Abhishek[3], V. Ramakrishna Sharma[4]
[1, 2, 3]Student, [4]Associate Professor, ECE, Sreenidhi Institute of Science and Technology, Ghatkesar,Hyderabad

Abstract: Music genre labels are useful to organize various songs, albums, and artists into broader groups that share related musical genres such as similar sound etc. A music genre is a conventional group it helps us to identify some pieces of music as belonging to a shared tradition or set of conventions. It is to be distinguished from musical form and musical style, although this terms can be used as viceversa. Music can be divided into genres in varying ways such as into popular music and art music,hip hop music or religious music and secular music. The artistic nature of music means that these classifications are often biased and notorious and some genres may overlap. We will classify the various music genres by using deep learning algorithm. We will train the model and by using various music genres of test dataset we will predict the specific music genre.

I. INTRODUCTION

Generation and classification is an important function with many real world applications. As the amount of music released daily continues to explode, especially on online platforms such as Sound cloud and Spotify number 2016 suggests that tens of thousands of songs are released every month on Spotify the need for precise metadata needed for data management and search or retention purposes. The ability to quickly distinguish songs from any playlist or library by the type of work that works for any music streaming and services as well as the analytical power of editing and completing the music and audio label is unlimited. We have used the genre of genre using our neighbours. We have used different methods of classification algorithms to separate the two types of input. We have tried the RBF kernel vector support machine the k nearest neighbours. The basic network for advanced feeds and finally the high level and advanced convolutional neural network. In addition to our algorithms we experimented with both raw amplitude data and modified mel spectrograms of that green amplitude data. We then release the predicted genre for the 10 most common genres of music. We have found that converting our green audio into mel spectrograms has produced better results for all of our models our convolutional neural network exceeds human accuracy. Mechanical learning methods have been used to classify the genre of music to determine now. In 2002 G. Tzanetaki Cook uses a mixture of Gaussians model and close neighbours k and three sets of carefully crafted objects representing timbral texture, rhythmic content and voice content.They achieved 61% accuracy. As a benchmark human accuracy is at least 70% of this type of sorting task Tzanetakis and Cook use MFCCs a close cousin of mel spectrograms and in fact all work follows in their footsteps in converting their data in this way. In subsequent years methods such as vector support systems were also used in this work as in 2003 when used multiple layers of SVMs to achieve more than 90% accuracy in a database containing only four genres. Over the past 5 to 10 years the convolutional neural networks have proven to be extremely accurate in the analysis of the genre, with positive results showing both the difficulties brought about by having multiple layers and the ability of the convolutional layers to accurately identify patterns within images (which are actually mel spectrograms and MFCCs).These results greatly exceeded the human capacity for genre classification with our research finding that modern models work with about 91% accuracy when using the full length of a 30s track. Many papers have used CNN to compare their models with other ML strategies, including k-NN, a mixture of Gaussians, and SVMs, and CNNs have worked well in all cases. We have therefore decided to focus our efforts on using the most accurate CNN, with other models being used as a starting point.

In this review paper we will analyze the music genre classification using knn algorithm.

II. LITERATURE REVIEW

The identification of musical genres has given a great deal of attention to various subject disciplines and a popular topic in MIR studies over the decades. The main and collaborative goal of a variety of methods is to classify music genres in a precise way as to provide an appropriate summary of music details. These methods based on a variety of methods cannot be listed but the study of Correa and Rodrigues examines studies from many years ago and reveals very different approaches. Depending on the problems of the division of Correa and Rodrigues can be dealt with according to the five different methods from the MIR community. This is audio based, graphic-based content, military-based, community-based meta data and hybrid methods. Audio based approach explores digital and audio signals to detect music features while content based view collects characterstics from different data forms that place music at a higher level. These content based approaches have attracted a lot of attention recently in MIR studies. Audio based system updates released by Fu and others. Music classification methods, an engaging discussion about key related subjects, their definitions based on music characters, classification schemes and performance descriptions. The review includes five major areas such as genre and air segregation, artist identification, instrument recognition and music annot ation. Song-based methods create a distinction based on musical knowledge as textual digging techniques are used. Lyric-based classification can be considered a problem of natural language processing. In this type of problem, the purpose is to provide labels and meaning to texts. That means separating the type of lyrical text in this crate. Similarly the methods for public meta data are aimed at obtaining online music information related to songs while benefiting from web mining techniques. Hybrid-based programs use characterstics from previous approaches. In this section of the paper type it is read in terms of content. A key step in a content-based approach is to view the entire audio signals. The purpose is to predict sub genres / songs while using compatib le music features that are automatically added to the sound. In addition to the content-based approach the point of pride of the paper gives importance to the use of songs. Examining specific sections and complete versions of songs separately provides a different method of sorting according to the characteristics of the songs. The Study of paper of Basil, Seraphim, and Stellato provides a fine example. Their work uses different machine learning algorithms to divide music genres into "more popular genres" based on trained examples (Basili et al., 2004). Various combinations of musical symbols are used to determine what can bring the most accurate results. This feature of crate music includes musical instruments, instrument classes, meter change and time and expansion note / width. Using these algorithms different algorithms are used to classify the types according to them and the details are compared at the end.

III. METHODS

A. Audio Retrieval

Audio retrieval is the interdisciplinary science of retrieving information from music. Audio retrieval is a small but growing field of research with many real world applications. Those involved in audio retrieval may have a background in musicology, psychoacoustics, psychology, academic music study, signal processing, informatics, deep learning, optical music recognition, computational intelligence or some combination of these. In Audio retrieval process we have retrieved the music from the youtube database. In the audio retrieval process the vedio files of different music genres is obtained from the youtube database. We have used the youtube_dl module to import the vedios from the youtube. We have converted the data into the music genres and loaded in the csv file by using pandas library. We have imported the various music files from the youtube because from the youtube, we can get the large collection of music files obtaining music files in the large number will allow us to get more accurate results than the preferred dataset used in various music genres that is gtzan dataset.

B. Spectrogram Visualization

Spectrogram is defined as the representation of spectrum of frequencies which varies with respective to time.In the spectrogram the signal is represented visually. Spectrograms are also called as the sonographs, voice prints or voicegrams. The audio data represented in the 3 dimensional graph are said to be the waterfalls. Spectrograms are widely used in radar for identification of the target object and identifying music genres which is performed by the audio signal processing. The spectrograms are also used in the seismology and various applications. The main use of Spectrograms is to identify the spoken words graphically.It is also used to identify the voices of the animals.The spectrogram is generated by the specific instruments commonly used 3 instrument is said to be optical spectrometer. Optical spectrometer uses a band pass filters which internally uses fourier transform or wavelet transform. Spectrograms are usually generated by using the discrete wave transforms or continuous wave transforms which are generated for the sample of audio signal. Spectrogram of audio signal represents the audio waves in form of intensities. The darker part in the spectrogram represents the higher frequency part and lighter part represents the low frequency part.

C. VGG Transfer Learning

Transfer learning is the problem which arised during research of deep learning it generally focuses on storing the patterns of the data and applying the data to the different but the problem is related to the model. For example knowledge gained while learning to recognize flowers could apply when trying to recognize plants. This area of research obtains some relation to the long history of psychological literature on transfer of learning although formal ties between the two fields are limited. From the pratical analysis we can say that the patterns obtained from one analysis is used to obtain the analysis of the other task.

In the year 1976 Stevo Bozinovski and Ante Fulgosi published a paper explicitly addressing transfer learning in neural networks training. The paper gives a complete demonstration of the mathematical and geometrical model of transfer learning. In the year 1981 a complete report was generated on the application of transfer learning in training a neural network it is performed on a dataset of images representing letters of computer terminals. Both positive and negative transfer learning was experimentally demonstrated in the year 1981. A paper was published in the year 1993 by Lorien Pratt the paper deals with the transfer in deep learning formulating the discriminability based transfer algorithm.

D. VGG Fine Tuning

The fine tuning is the process of varying parameters of a model must be in such a way that it fits with certain observations. The fine tuning had led to the discovery of the fundamental constants and quantities fall extraordinarily precise range that if fine tuning did not the origin and evolution of agents who are conscious in the universe would not be permitted. The Theories which require fine tuning are regarded as problematic in the absence of a actual mechanism to explain why the parameters are precisely happen and the values which are observed are returned. Naturalness is defined as the heuristic rule the parameters in a fundamental physical theory should not be fine tuned. The idea of naturalness concept which explain fine tuning was brought into question by Nima Arkani Hamed a theoretical physicist in the lecture of talk why is there a macroscopic universe which exists in the miniseries Multiverse and Fine Tuning from the Philosophy of Cosmology project which is collaborated by the University of Oxford and Cambridge Collaboration 2013.

E. VGG Feed Forward Baseline Model

The feed forward Neural Network is an essential part of the artificial neural network to the connectionist learning, a procedure known as the method of control, which requires that the teacher, in order to display the desired vector. FNN is one of the parts of a multi-layer perceptron with one-way flow of information. These are the neurons are arranged in layers, namely, input, hidden, and output layers, and the connections between the neurons in a layer, and the neurons in the next layer . In this regard, it is repeated twice to be adjusted in order to minimize the the the the the best, the difference between the rest, the direction of the network, and to select the rest, the direction of the learning process. There are usually trained with the help of an algorithm in the reverse error propagation as the use of a learning curve, which is called gradient descent, which is also known as steepest descent. In this paper we will present the output of music genres using music genre classification

IV. METHODOLOGY

A. Existing Methodology

In the paper of Music genre classification of audio signals proposed by George tanzatakis and perry cook, the 10 genres of music was classified as three important features timbral texture, rhythmic content and pitch content. The 10 musical genres are classified by machine learning algorithms. The accuracy achieved by them was 61%.

B. Proposed Methodology

In this paper we have used the gtzan dataset which contains the 1000 music files in the form of 10 music genres. We have performed the classification of music genres by using knn algorithm which is nothing but the machine learning algorithm. The accuracy of our proposed methodology is 78%.Our methodology has high accuracy when compared to the method proposed in the music genre classification of audio signals. In the existing methodology the statistical pattern recognition classifier was used and accuracy obtained by them was 61%.

In our proposed methodology we have implemented by using vgg 19 model. We have performed the three stages. Vgg 19 transfer learning validation accuracy obtained was 75%. Vgg 19 fine tuning validationaccuracy obtained was 78%.The final stage of methodology was feed forward baseline the accuracy obtained was 81%.

V. RESULTS

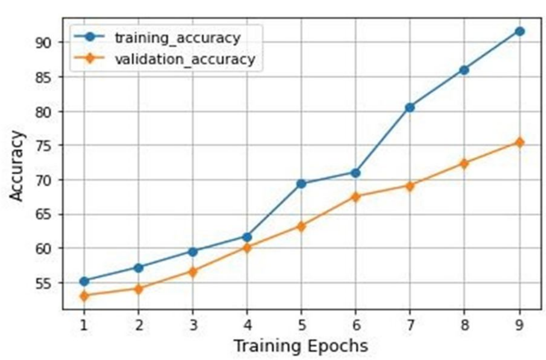

Fig 1.Training and validation accuracy of VGG transfer learning

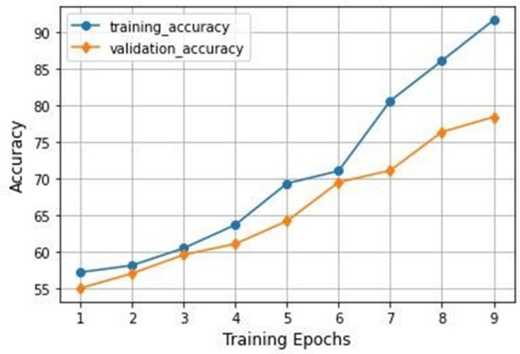

Fig 2.Training and validation accuracy of VGG fine Tuning

Fig 3.Training and validation accuracy of VGG Feed Forward Baseline

VI. CONCLUSION

We have done the music genre classification using CNN deep learning Technique. We have trained and tested the data. By this CNN Technique we have got the accuracy 85.7%. The accuracy can be improved by using the various other algorithms. Motivated by the interest of deep learning in the Music Genre Classification task, we decided to use a map of eight musical features as inputs of a CNN.

These features were chosen along dynamics, timbre and tonality dimensions, among a larger set studied in an early work. We relied on CNNs trained in such a way that filter dimensions (adapted for our purpose) are interpretable in time and frequency. Results show the relevance of our eight music features: global accuracy of 85.7% against 77.8% for 513 frequency bins of a spectrogram. The late score fusion between systems based on both feature types reaches 86 % accuracy on the GTZAN database.

VII. REFRENCES

[1] Ossama Abdel Hamid, Abdel rahman Mohamed, HuiJiang, Li Deng, Gerald Penn, and Dong Yu. 2014.Convolutional neural networks for speech recognition. IEEE/ACM Transactions on audio, speech, and language processing 22(10):1533– 1545.
[2] Yali Amit and Donald Geman. 1997. Shape quantization and recognition with randomized trees. Neural computation 9(7):1545–1588.
[3] Leo Breiman. 1996. Bagging predictors. Machine learning 24(2):123–140.
[4] Leo Breiman. 2001. Random forests. Machine learning 45(1):5–32.
[5] Corinna Cortes and Vladimir Vapnik. 1995. Supportvector networks. Machine learning 20(3):273–297.
[6] Steven B Davis and Paul Mermelstein. 1990. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. In Readings in speech recognition, Elsevier, pages 65–74.
[7] Dan Ellis. 2007. Chroma feature analysis and synthesis. Resources of Laboratory for the Recognition and Organization of Speech and Audio-LabROSA.
[8] Jerome H Friedman. 2001. Greedy function approximation: a gradient boosting machine. Annals of statistics pages 1189–1232.
[9] Jort F Gemmeke, Daniel PW Ellis, Dylan Freedman, Aren Jansen, Wade Lawrence, R Channing Moore, Manoj Plakal, and Marvin Ritter. 2017. Audio set: An ontology and human-labeled dataset for audio events. In Acoustics, Speech and Signal Processing(ICASSP), 2017 IEEE International Conference on IEEE, pages 776–780.
[10] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard,
[11] Hareesh Bahuleyan. Music genre classification using machine learning techniques. CoRR, abs/1804.01149, 2018.