© 2018 IJSRSET | Volume 4 | Issue 4 | Print ISSN: 2395-1990 | Online ISSN : 2394-4099
Themed Section : Engineering and Technology

Music Genre Classification using Deep Neural Networks

G. Jawaherlalnehru*[1],S. Jothilakshmi[2]
[1]Research Scholar Department of Computer Science & Engineering, Annamalai University, Chidambaram, Tamil Nadu, India
[2]Assistant Professor Department of Information Technology, Annamalai University, Chidambaram, Tamil Nadu, India

ABSTRACT

genres can be defined as categorical labels created by humans to identify or characterize the style of music. This work presents a comprehensive machine learning approach to the problem of automatic musical genre classification using the audio signal. The system is developed using a Deep Neural Network (DNN) to recognize the genres. Mel Frequency Cepstral Coefficients (MFCC) features are used to represent the music characteristics. The system is evaluated with MIR datasets. The proposed system observed higher classification accuracy of 97.8%.

Keywords: Music Information Retrieval, Music genre classification, Deep learning, Mel frequency cepstral coefficients.

I. INTRODUCTION

Music, songs, and videos are abundant on the web and keep growing. The large scale music files depend on automated algorithms to analyze and index music content. Hence automatically extracting music information is gaining importance as a way to organize and structure music files. Music information retrieval (MIR) is an emerging research area to cope with such necessity. The MIR tasks include artist identification, genre classification, emotion recognition, cover song recognition, instrument recognition, and music annotations. Feature extraction and classifier learning are the key components of music information retrieval. With the growth of the internet, multimedia systems applications that deal with the musical databases gained importance and demand for Music Information Retrieval (MIR) applications increased. Automatic analysis of the musical databases is one of the required components of the MIR [1]. Music genre classification is a widely studied problem in the Music Information Research (MIR) community. However, almost all related work is concentrated in multiclass classification of music items into broad genres (e.g., Classical, Electronic, Pop, Rock), assigning a single label per item [2].

Musical genres are categorical labels created by human experts and it is used for categorizing, describing and even comparing songs, albums, or authors in the vast universe of music [4]. Artificial neural networks have great potential in learning complex high level knowledge from raw inputs thanks to its non-linear representation of the hypothesis function. Training and generalizing a neural network with many hidden layers using standard techniques are challenging [11]. Most of the current music databases are indexed based on song title or artist name and in this format improper indexing can result in incorrect search results [1]. Previous studies that deal with genre classification use symbolic representations of music such as MIDI files [3, 5, 6] or wav or mp3 [7, 8] audio files. Most of the proposed methods have two processing steps. The first one is frame-based feature extraction step where feature vectors are computed from each frame. Most

common features used for genre classification are; timbre, rhythm and pitch-related features [9]. Timbre related features are; Fast Fourier Transform coefficients, Cepstrum and Mel Frequency Cepstrum Coefficients (MFCC), Linear Prediction (LP) coefficients, MPEG filter bank components, Spectral Centroid, Spectral Flux, Zero Crossing Rate, Spectral Roll-Off, low order statistics and Delta coefficients [7]. More detailed descriptions of the features can be found in [7, 9, 10]. In the second step pattern recognition algorithms are applied on the feature vectors to achieve genre classification [7].

Recently, deep learning is becoming more popular for supervised learning applications such as environment sound classification, and speech recognition. Specifically we would like to apply various deep learning algorithms to classify music genres and study their performances. More recent deep learning approaches take advantage of visual representations of the audio signal in form of spectrograms. Many recent findings show that deep neural networks in some tasks outperform most of the traditional classification algorithms such as support vector machine and random forest [11]. In this work deep learning algorithm are applied to classify music genres and analyzed its performance.

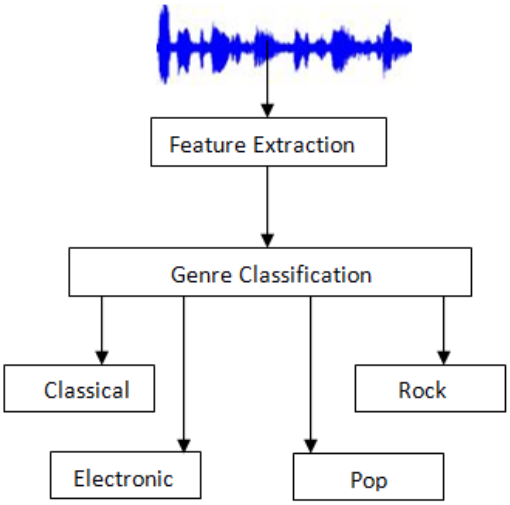

Fig.1 Automatic Genre Classification

The rest of the paper is structured as follows: A brief overview on related work is provided in Section II. Section III deals with Dataset details and feature extraction methods. The proposed Deep Neural Network (DNN) based automatic music genre classification is presented in Section IV. Section V provides conclusions and suggests potential directions for future improvements

II. RELATED WORK

Automatic musical genre identification does not have a long history but there has been a lot of interest in the recent years. It is quite interdisciplinary and draws especially from areas such as digital signal processing, machine learning, and music theory. Most published music genre classification approaches rely on audio sources [2, 18]. Traditional techniques typically use handcrafted audio features, such as Mel Frequency Cepstral Coefficients (MFCCs) [12], as input of a machine learning classifier (e.g., SVM) [13, 14]. One of the most significant proposals specifically to deal with studies on automatic musical genre classification was proposed by Tzanetakis and Cook in 2002 [19]. In this work, the researchers use timbral related features, texture features, pitch related features based on the multi-pitch detection algorithm and the rhythmic content features based on Beat Histogram. For classification and evaluation, the authors used Gaussian Mixture Model (GMM) and k Nearest Neghbor (kNN) classifiers. The overall genre classification accuracy of the system reaches a 61% of correct classifications over the 10 musical genres. Sigtia and Dixon [20] trained a neural network with rectified linear units (ReLUs) using Stochastic Gradient Descent (SGD). They used the activations of the hidden layers of the neural networks as features and trained a Random Forest classifier on top of these features to predict the classes. The authors validated their ideas on two datasets, GTZAN and ISMIR 2004. In the former, they reported an accuracy of 83% while in the latter the best performance was74.4% of recognition rate. In [21], a music genre classification system was developed with various genres of Indian music. The authors had inferred that the combination of features such as MFCC, Spectral Centroid,

Skewness, Kurtosis, Flatness, Entropy, and Irregularity achieves classification accuracy of 91.25%, with Gaussian Mixture Model classifier.

III. DATASET AND FEATURE EXTRACION

3.1 Data Collection

In this work four music genres such as classical, pop, rock, and electronic are considered. The music database consists of 400 audio tracks with metadata. For each genre 100 audio tracks of 60sec long are considered. All the audio files are in .au format with 44.1 KHz sampling frequency, stereo and 16bit PCM.

3.2 Feature Extraction

Audio features act as a quantitative way to provide the most important information in an audio file. The process of extracting relevant characteristics enclosed within the input data is called as feature extraction. This process converts an audio signal into a sequence of feature vectors. Feature extraction reduces the redundant information from audio signal and provides a compact representation. Audio features can be divided into two levels as top-level and low-level according to perspective of music understanding [16]. The top level labels provide information on how listeners interpret and understand music using different genres, moods, instruments, etc. Low-level audio features can also be categorized into short-term and long term features on the basis of their time scale [17]. Some of audio features are temporal domain features frequency domain features, modulation frequency domain features. There are many feature extraction techniques such as Linear Prediction (LPC) Analysis, Spectral Centroid, Spectral Flux, Zero Crossing Rate (ZCR), Linear Predictive Cepstral Coefficients (LPCC), Fast Fourier Transform (FFT), Mel scale Cepstral analysis (MEL), Spectral Roll-Off, low order statistics and Delta coefficients. The most commonly used in speech recognition are the MFCCs. MFCC are one of the most popular set of features used in pattern recognition. MFCC was originally developed for automatic speech recognition systems, lately have been used with success in various musical information retrieval tasks [19]. MFCC takes human perception sensitivity with respect to frequencies into consideration, and therefore are best for speech/speaker recognition. MFCC are compact, short time descriptors of the spectral envelope audio feature set and typically computed for audio segments of 10-100ms. MFCC is a short term power spectrum of sound. It is based on a linear cosine transform a log power spectrum on a nonlinear mel scale of frequency. Each short term Fourier transform magnitude coefficient is multiplied by the corresponding filter gain and the results are accumulated. Then discrete cosine transform is applied to the log of the mel spectral coefficients to obtain MFCC. Fig.2 illustrates various steps involved in extracting MFCC features from raw audio signal.

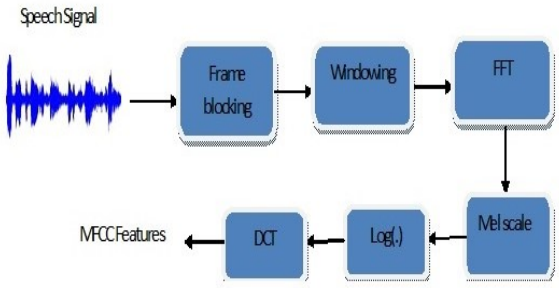

Figure 2. MFCC feature extraction

IV. PROPOSED MUSIC GENRE CLASSIFICATION SYSTEM

After the feature extraction process it is important to classify the input audio signal. Classification is the must process by which particular label is assigned to a particular music format. The classifier denotes decision boundaries in the feature space, which separates different sample classes from each other. Multi label feed-forward Deep Neural Network (DNN) architecture to various feature representations are generated using signal processing methods. Deep learning is a branch of machine learning based on a set of algorithms that attempt to model high-level abstractions in data by using model architectures, with complex structures, composed of multiple non-linear transformations [15, 16]. Various deep learning

architectures such as deep neural networks, convolutional deep neural networks, deep belief networks, and recurrent neural networks have been applied to fields like computer vision, automatic speech recognition, natural language processing, audio recognition, and bioinformatics where they have been shown to produce state-of-the-art results on various tasks [16, 17]. A deep-learning network trained on labeled data can then be applied to unstructured data, giving it access to much more input than machine-learning nets. The more data a net can train on, the more accurate it is likely to be. Deep-learning networks perform automatic feature extraction without human intervention, unlike most traditional machine-learning algorithms.

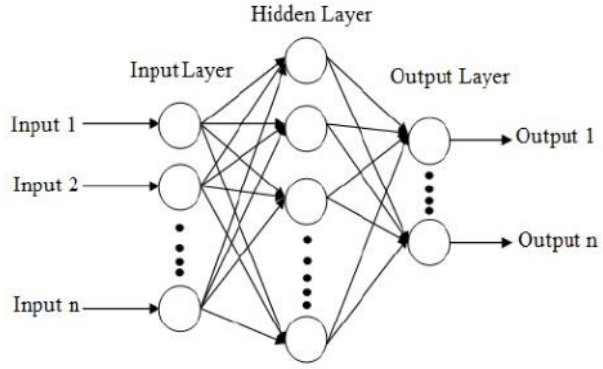

Figure 3. Deep Neural Network Architecture

Fig. 3 shows a multi-layer perceptron network of three layers. DNN is a non- linear multi-layer model with powerful capability to extract relevant features related to classification. The multilayer perceptron (MLP), also called Feed Forward Network, is the most typical neural network model. The biggest advantage of DNN is to extract and learn features automatically by deep layer architecture, especially for a complex and high dimensional data. The DNN used in this work is fully connected neural network which consists of one input layer, one output layer, and several hidden layers. The number of neurons of input depends on the dimensions of input feature vectors, while the number of neurons of output layers is equal to the number of music genres being considered. The inputs and outputs of the model's units follow the basic logic of the single neuron. The basic unit in this model is a neuron which is shown in Fig. 4.

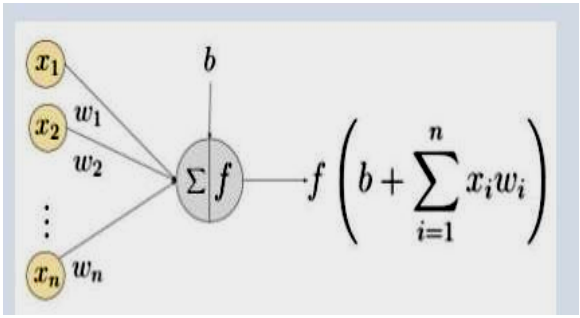

Figure 4 Basic unit of neuron

V. EXPERIMENTS AND RESULTS

In this work, R machine learning library 'H2O' Deep Learning is used to implement DNN. H2O is fast, scalable, open-source machine learning and deep learning for Smarter Applications. Advanced algorithms, like Deep Learning, Boosting, and Bagging Ensembles are built-in to help application designers create smarter applications through effective Application Programming Interfaces. For this work, the audio segments are converted to 16 bit mono. But the original sampling frequency is retained without down sampling. The implementation of deep learning architecture uses log mel band energy features. For classification of music genres, the dataset is partitioned randomly into three parts: 60% for training, 20% for validation, 20% for testing. This model is evaluated using four cross fold validation scheme. For each, per-class accuracy is calculated on frame-wise level. These scores are obtained by dividing number of correctly predicted frames by the total number frames belonging to that class. Finally, the overall score is calculated by averaging the cross fold accuracy. The performance of DNN is evaluated by varying the number of epochs up to 200.

In this experiment uses hidden layers is 350, and epochs is 60 it's followed by rectifier activation. The accuracy is 97.8%. The results are shown in confusion matrix table 1, and the overall performance is given in the figure 5.

Table 1: Confusion matrix of proposed work for each class

	classical	electronic	pop	rock	Error	Rate
classical	2268	148	0	0	0.06125828	148 / 2,416
electronic	0	2455	40	0	0.01603206	40 / 2,495
pop	0	0	2347	168	0.06679920	168 / 2,515
rock	0	0	0	2508	0.00000000	0 / 2,508
Totals	2268	2603	2387	2676	0.03583652	356 / 9,934

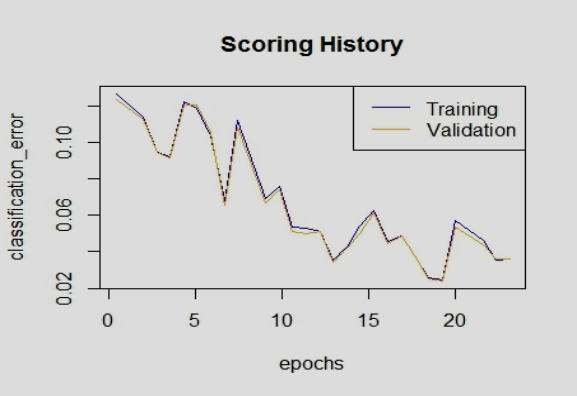

Figure 5. Shows on Plot diagram for overall Performance

V. CONCLUSION AND FUTURE WORK

In this work, a deep learning approach is proposed for automatic genre identification and it is evaluated with GENRE Dataset. DNN with MFCC features works well better than the other system. The accuracy is 97.8% in the proposed system. DNN is the best classifier model better than the other classifier model. The future work includes with increasing databases with other feature techniques in such a way that the classification task. We intend to experiment bigger amounts of data (Million Song Dataset) and develop new features which are able to extract the musically-meaningful information from the audio signals and use more feature sets such as melodic characteristics. This should be an effective method to combine ensemble method with deep neural networks architecture.

VI. REFERENCES

[1]. Yusuf Yaslan and Zehra Cataltepe "Audio Music Genre Classification Using Different Classifiers and Feature Selection Methods", 2006.

[2]. Dmitry Bogdanov, Alastair Porter, Perfecto Herrera, and Xavier Serra. Cross-collection evaluation for music classification tasks. In ISMIR, 2016.

[3]. Sergio Oramas, Oriol Nieto, Francesco Barbieri, Xavier Serra "Multi-Label Music Genre Classification From Audio, Text, And Images Using Deep Features", ISMIR 2017.

[4]. Dhanith Chathuranga, Lakshman Jayaratne "Automatic Music Genre Classification of Audio Signals with Machine Learning Approaches", GSTF Journal on Computing (JoC), August 2013.

[5]. Tao Feng "Deep learning for music genre classification"courses.engr.illinois.edu. 2014.

[6]. Beth Logan et al. Mel frequency cepstral coefficients for music modeling. In ISMIR, 2000.

[7]. Klaus Seyerlehner, Markus Schedl, Tim Pohle, and Peter Knees. Using block-level features for genre classification, tag classification and music similarity estimation. Submission to Audio Music Similarity and Retrieval Task of MIREX, 2010.

[8]. George Tzanetakis and Perry Cook. Musical genre classification of audio signals. IEEE Transactions on speech and audio processing, 10(5):293–302, 2002.

[9]. Betsy. S, D. G. Bhalke "Genre Classification of Indian Tamil Music using Mel-Frequency Cepstral Coefficients" , International Journal of Engineering Research & Technology (IJERT), Vol. 4 Issue 12, December-2015.

[10]. Rosner, Aldona; Schuller, Bjorn; Kostek, Bozena, "Classification of Music Genre Based on Music Separation into Harmonic and Drum Components", Archives of Acoustics S.Iv.39, n.4, p.629-638, dec 2014.

[11]. Tzanetakis, G.; Cook, P., "Musical genre classification of audio signals, "IEEE Transactions on Speech and Audio Processing, vol.10, no.5, pp.293-302, Jul 2002.

[12]. Shih-Hao Chen; Shi-Huang Chen; Trieu-Kien Truong, "Automatic music genre classification based on wavelet package transform and best basis algorithm," IEEE International Symposium on Circuits and Systems (ISCAS), 2012.

[13]. Meng, A.; Ahrendt, P.; Larsen, J.; Hansen, L.K., "Temporal Feature Integration for Music Genre Classification", IEEE Transactions in Audio, Speech, and Language Processing, vol.15, no.5, pp.1654-1664, July 2007.

[14]. Scaringella, N.; Zoia, G.; Mlynek, D., "Automatic genre classification of music content: a survey," IEEE Signal Processing Magazine, vol.23, no.2, pp.133,141, March 2006.

[15]. Y Bengio, Learning Deep Architectures for AI. (2009),http://www.iro.umontreal.ca/~bengioy/papers/ftml.pdf. Accessed 26 Jan 2016

[16]. AB Olshausen, Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature 381(6583), 607–609 (1996).

[17]. L Gome, Machine-Learning Maestro Michael Jordan on the Delusions of Big Data and Other Huge Engineering Efforts.(2014),http://spectrum.ieee.org/robotics/artificialintelligence/machinelearning-maestro-michael-jordan-onthe-delusions-of-big-data-and-other-huge-engineering-efforts. Accessed 26 Jan 2016.

[18]. Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1–9, 2015.

[19]. Logan, B. Mel Frequency Cepstral Coefficients for music modeling. Read at the first International Symposium on Music Information Retrieval, 2000.

[20]. S. Sigtia, S. Dixon, Improved music feature learning with deep neural networks, in: IEEE International Conference on Acoustic, Speech and Signal Processing, 2014, pp. 6959–6963.

[21]. S Jothilakshmi, N. Kathiresan "Automatic Music Genre Classification for Indian Music" IPCSIT vol. 41, 2012.