Music Genre Classification using MFCC and AANN

R. Thiruvengatanadhan[1]

[1]*Assistant Professor/Lecturer (on Deputation), Department of Computer Science and Engineering, Annamalai University, Annamalainagar, Tamil Nadu, India*

Abstract *- Musical genres are categorical descriptions that are used to describe music. They are commonly used to structure theincreasing amounts of music available in digital form on the Web and are important for music information retrieval. Music classification serves as the fundamental step towards the rapid growth and useful in music indexing. Searching and organizing are the main characteristics of the music classification system these days. This paper describes a technique that uses Auto associative neural network (AANN) to classify songs based on features using Mel Frequency Cepstral Coefficients (MFCC). Experimental results of multi-layer support vector machines shows good performance in musical classification.*

Key Words: Feature Extraction, *Mel Frequency Cepstral Coefficients* (*MFCC)* and *Auto associative neural network (AANN)*.

1. INTRODUCTION

Musical genres are categorical descriptions that are used to characterize music in music stores, radio stations and now on the Internet. Although the division of music into genres is somewhat subjective and arbitrary there are perceptual criteria related to the texture, instrumentation and rhythmic structure of music that can be used to characterize a particular genre. Humans are remarkably good at genre classification as investigated in [1] where it is shown that humans can accurately predict a musical genre based on 250 milliseconds of audio. This finding suggests that humans can judge genre using only the musical surface without constructing any higher level theoretic descriptions as has been argued.

Therefore techniques for automatic genre classification would be a valuable addition to the development of audio information retrieval systems for music. Advanced music databases are continuously achieving reputation in relations to specialized archives and private sound collections. Due to improvements in internet services and network bandwidth there is also an increase in number of people involving with the audio libraries.

But with large music database the warehouses require an exhausting and time consuming work, particularly when categorizing audio genre manually. Music has also been divided into Genres and sub genres not only on the basis on music but also on the lyrics as well [2]. This makes classification harder. To make things more complicate the definition of music genre may have very well changed over time [3]. For instance, rock songs that were made fifty years ago are different from the rock songs we have today. Luckily, the progress in music data and music recovery has considerable growth in past years.

2. ACOUSTIC FEATURES FOR AUDIO CLASSIFICATION

An important objective of extracting the features is to compress the music signal to a vector that is representative of the meaningful information it is trying to characterize. In these works, acoustic features namely MFCC features are extracted.

2.1 Mel Frequency Cepstral Coefficients

Mel Frequency Cepstral Coefficients (MFCCs) are short-term spectral based and dominant features and are widely used in the area of audio and speech processing. The mel frequency cepstrum has proven to be highly effective in recognizing the structure of music signals and in modeling the subjective pitch and frequency content of audio signals [4]. The MFCCs have been applied in a range of audio mining tasks, and have shown good performance compared to other features.

MFCCs are computed by various authors in different methods. It computes the cepstral coefficients along with delta cepstral energy and power spectrum deviation which results in 26 dimensional features. The low order MFCCs contains information of the slowly changing spectral envelope while the higher order MFCCs explains the fast variations of the envelope [5].

MFCCs are based on the known variation of the human ears critical bandwidths with frequency [6]. The filters are spaced linearly at low frequencies and logarithmically at high frequencies to capture the phonetically important characteristics of speech and audio. To obtain MFCCs, the audio signals are segmented and windowed into short frames of 20 ms.

Magnitude spectrum is computed for each of these frames using Fast Fourier Transform (FFT) and converted into a set of mel scale filter bank outputs. The human ear resolves frequencies non-linearly across the audio spectrum and empirical evidence suggests that designing a front-end to operate in a similar non-linear manner improves the performance. A popular solution is therefore filterbank analysis since this provides a much more straightforward route to obtain the desired non-linear frequency resolution. However, filterbank amplitudes are highly correlated and hence, the use of a cepstral transformation in this case is virtually mandatory. Fig. 1 describes the procedure for extracting the MFCC features.

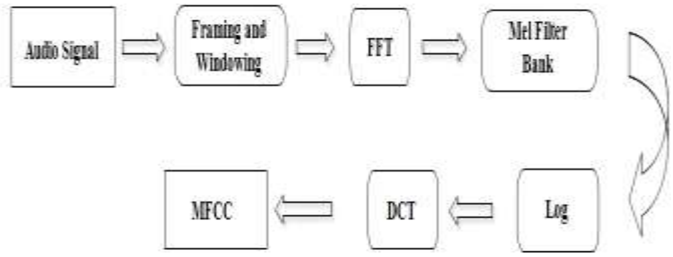

Fig -1: Extraction of MFCC from Audio Signal.

Mel frequency to implement this filterbank, the window of audio data is transformed using a Fourier transform and the magnitude is taken. The magnitude coefficients are then binned by correlating them with each triangular filter. Here, binning means that each FFT magnitude coefficient is multiplied by the corresponding filter gain and the results are accumulated. Thus, each bin holds a weighted sum representing the spectral magnitude in that filterbank channel.

Logarithm is then applied to the filter bank outputs. Discrete Cosine Transformation (DCT) is applied to obtain the MFCCs. Since the mel spectrum coefficients are real numbers, they are converted to the time domain using the DCT.

In practice, the last step of taking inverse Discrete Fourier Transform (DFT) is replaced by taking DCT for computational efficiency. The cepstral representation of the speech spectrum provides a good representation of the local spectral properties of the signal for the given frame analysis. Typically, the first 13 MFCCs are used as features.

3. CLASSIFICATION MODEL

3.1 Autoassociative Neural Network (AANN)

Autoassociative Neural Network (AANN) model consists of five layer network which captures the distribution of the feature vector as shown in Fig. 2. The input layer in the network has less number of units than the second and the fourth layers. The first and the fifth layers have more number of units than the third layer [7]. The number of processing units in the second layer can be either linear or non-linear. But the processing units in the first and third layer are non-linear. Back propagation algorithm is used to train the network [8].

The activation functions at the second, third and fourth layer are nonlinear. The structure of the AANN model used in our study is 13L 26N 4N 26N 13L for capturing the distribution of acoustic features, where L denotes a linear unit, and N denotes anon-linear unit. The integer value indicates the number of units used in that layer [9]. The non-linear units use tanh(s) as the activation function, where s is the activation value of the unit. Back propagation learning algorithm is used to adjust the weights of the network to minimize the mean square error for each feature vector [10].

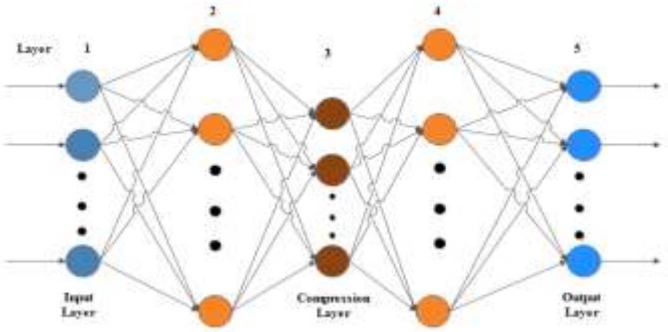

Fig -2: Auto associate neural network.

4. EXPERIMENTAL RESULTS

4.1 Dataset Collection

The music data is collected from music channels using a TV tuner card. A total dataset of 100 different songs is recorded, which is sampled at 22 kHz and encoded by 16-bit. In order to make training results statistically significant, training data should be sufficient and cover various genres of music.

4.2 Feature Extraction

In this work fixed length frames with duration of 20 ms and 50 percentages overlap (i.e., 10 ms) are used. The objective of overlapping neighboring frames is to consider the temporal characteristic of audio content. An input wav file is given to the feature extraction techniques. MFCC 13 dimensional feature values will be calculated for the given wav file. The above process is continued for 100 number of wav files.

4.3 Classification

When the feature extraction process is done the music should be classified. We select 75 music samples as training data including 25 classic music, 25 pop music and 25 rock music. The rest 25 samples are used as a test set. The feature vectors are given as input and compared with the output to calculate the error. In this experiment the network is trained for 500 epochs. The confidence score is calculated from the normalized squared error and the category is decided based on highest confidence score. The network structures 13L 26N 4N 26N 13L gives a good performance and this structure is obtained after some trial and error.

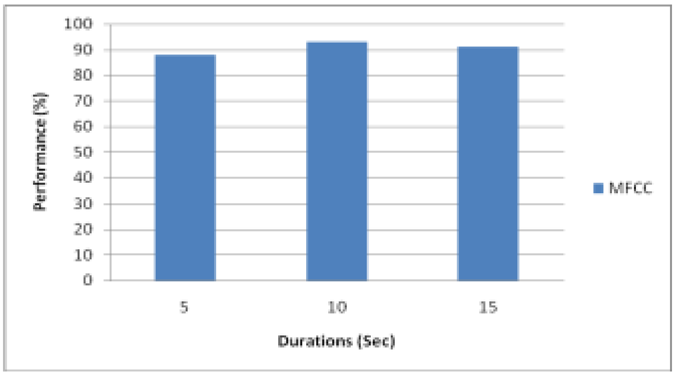

Chart -1: Performance of music genre classification for different duration of music clips

The performance of AANN for different duration as shown in Chart 1 shows that when the duration were increased from 10 to 15 there was no considerable increase in the performance.

5. CONCLUSION

In this paper, we have proposed an automatic music classification system using AANN. MFCC is calculated as features to characterize audio content. AANN learning algorithm has been used for the classification of genre classes of music by learning from training data. The confidence score is calculated from the normalized squared error and the category is decided based on highest confidence score between classic and pop, pop and rock by learning from training data. Experimental results show that the proposed audio AANN learning method has good performance in musical genre classification scheme is very effective and the accuracy rate is 93%.

REFERENCES

[1] Perrot, D., and Gjerdigen, R.O. Scanning the dial: Anexploration of factors in the identification of musical style. In Proceedings of the 1999 Society for Music Perception and Cognition pp.88

[2] Serwach, M., & Stasiak, B. (2016). GA-based parameterization and feature selection for automatic music genre recognition. In Proceedings of 2016 17th International Conference Computational Problems of Electrical Engineering, CPEE 2016.

[3] Dijk, L. Van. (2014). Radboud Universiteit Nijmegen Bachelorthesis Information Science Finding musical genre similarity using machine learning techniques, 1–25.

[4] O.M. Mubarak, E. Ambikai rajah and J. Epps, "Novel Features for Effective Speech and Music Discrimination," IEEE Engineering on Intelligent Systems, pp. 342-346, 2006.

[5] A. Meng and J. Shawe-Taylor, "An Investigation of Feature Models for Music Genre Classification using the Support Vector Classifier," International Conference on Music Information Retrieval, Queen Mary, University of London, UK, pp. 604-609, 2005.

[6] A. Dessein and A. Cont, "An Information-Geometric Approach to Real-Time Audio Segmentation," Signal Processing Letters, IEEE, vol. 20, no. 4, pp. 331-334, 2013.

[7] ShaojunRen, Fengqi Si, Jianxin Zhou, Zongliang Qiao, Yuanlin Cheng, "A new reconstruction-based auto-associative neural network for fault diagnosis in nonlinear systems," Chemometrics and Intelligent Laboratory Systems, Volume 172, 15 January 2018, Pages 118-128N.

[8] Nitananda, M. Haseyama, and H. Kitajima, "Accurate Audio-Segment Classification using Feature Extraction Matrix," IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 261-264, 2005.

[9] G. Peeters, "A Large Set of Audio Features for Sound Description," Technical representation, IRCAM, 2004.

[10] K. Lee, "Identifying Cover Songs from Audio using Harmonic Representation," International Symposium on Music Information Retrieval, 2006.