STM JOURNALS

Journal of Artificial Intelligence Research & Advances
ISSN: 2395-6720(online)
Volume 2, Issue 2
www.stmjournals.com

Identification of individual ragas using artificial neural network classifier

Priya, Ravi Kumar*
Electronics and Comm. Engineering Department, Thapar University, Patiala 147004, India

Abstract
This paper presents the feature extraction as the base step of music information retrieval. Basic feature of music is chroma, which is extracted using from the music signal using spectrogram and chromagram. After the extraction of feature the comparison is made between the original data and extracted chroma data using neural network.

Keywords: *Music information retrieval, artificial neural network, chromagram*

****Author for Correspondence*** E-mail: ravi.kumar@thapar.edu

INTRODUCTION
Music has taken up a very important place in the life of billions of people. People of all ages feel connected to one or other form of music. Local music and performance on music represents the culture and emotions of people of different areas. New trends in music distribution and storage have led to tremendous interest in the processing of music signals. From surfing music collections, to discovery new performers, and preventing the records from privacy computers have played a very vital role, thus processing of music is need of the hour[1]. Every part of music can be classified on the various basis- it can be genre, artist or many more aspects. The music samples have wide variety of information and details which are very important for different kind of applications. We can check various properties of audio signals such as fundamental periodicity- known as pitch, and amount of overlapping of music instruments in a sample- known as polyphony, and types of characteristics- known as timbre, and also on the basis of temporal structures- known as beats.[2] Typically, we associate music signal processing with speech signal analysis. However, the methodologies adopted are quite different in both the cases. Music signal processing poses more challenges as compared to speech processing since, the terms like, ‘melody’ and ‘genre’ have little statistical significance. Thus, it is often difficult to extract relevant features from the raw data for melody classification [3][4]. There are many features using which we can classify the samples. In this paper we classify on the basis of Chroma features and then training and testing using neural network.

FEATURE EXTRACTION
Music Data Base
In classical music theory, a specific music scale is used to write the piece of music. The scale defines the set of all the possible notes in the sample taken. Western music is defined as the list of notes, and details of the note like, its pitch and length. The pitch and length are considered to be the automatic units. While defining note information the combined information of note pitch and note lengths are defined [5]. The music which we have considered as our database for our experiment was 64 classical songs from various ragas.

Chroma Features
Most of the musical instruments, such as flutes, trumpets, violins, guitar, and pianos etc, are all based on the feature that helps musicians to play sound with great control and stable fundamental periods. Such types of signals are harmonic series of sinusoids at multiples of fundamental frequencies and the results in a percept of musical note. These notes have clear pitch in the listener’s mind. Music brings pleasure to human minds thus its features are ought to represent human auditory perception. In human perception two signals with fundamental frequencies fall in ratio of 2:1(octave) are highly similar. Sometimes this

is known as "octave equivalence". The sinusoidal harmonics of fundamental frequencies f0, 2f0, 3f0… are proper superset of the harmonics of note f0 with fundamental 2f0. In western major scale spans the octave twelve steps with basic notes- C, D, E, F, G, A, B. These notes are called 'white notes'. The space between two consecutive notes is two semitones apart from each other, except for E-F and B-C, these are only one semitone apart. These semitones are also referred to as 'black notes'. The semitone below note is generally referred to as #-sharp of that note like C-sharp (C#). The Octave degree is used to denote the Pitch's Chroma. Chroma and octave number when concatenated outcome is specific pitch. On a piano lowest note is A0 (27.5 Hz), middle is C4 (262 Hz) and highest is C8 (4186 Hz) [1][6][2].

Chromagram

The chromagram, also referred to as Harmonic Pitch Class profile, is the display or description of energy distributions along the pitch classes. Firstly, we need to define the chroma spectrum, X(n), it is the measure of signal's strength with respect to a given value of chroma C. Chroma C, is just the fractional part of the based 2 logarithmic output of frequencies, that is concluded with using Shepard's helix of pitch perception[6][7]

$$C = \log_2 f - \lceil \log_2 f \rceil \quad (1)$$

In equation (1), ⌈ ⌉ represents the greater integer function. The chroma spectrum represents the standard Fourier power spectrum. On frame fragmenting before calculating fourier spectrum, we can also create time-frequency distribution, that is denoted by X(t, f). We can define a "time-chroma" distribution, X(t ,C) as

$$X(t, C) = \text{fun}(X(t, f)) \quad (2)$$

Here, fun is an aggregation function, which is used in the remapping of the time frequency distribution. The distribution is named as 'Chromagram'. 'f' is given by 2^{C+n} which is obtained from equation (1) and the aggregation function we use here is summing function. Elements of chroma features for t^{th} frame of V (t, k) can be obtained by using the formula

$$V(t, k) = \sum_n \left(\frac{X_t(n)}{N_k}\right) \quad (3)$$

Where k ϵ {0, 1, 2, …, 11}
n ϵ S_k

$X_t(n)$ is the logarithmic result of the DFT (Discrete Fourier Transform) of the t^{th} frame. N_k is number of elements in S_k, which is subset of the discrete frequency region for each pitch class.

The arithmetic mean of log magnitude DFT values belonging to set S_k is taken and then each feature vector is normalized by subtracting scalar mean of the vector's feature set. The set S_k has 12 features generated by association of each DFT bin corresponding to each of the 12 pitch classes. The associated frequency of DFT bin is calculated and chroma value is obtained using formula (1).

The result saved the bin belongs to the pitch class which has the nearest chroma values. The chroma values are reassigned such that pitch class B has centre at chroma values 1, pitch class C has centre around value 0 an other pitch classes have their centre at chroma values k/12[8]. The range of spectrum is limited. The lower bound is chosen at 20 Hz and upper bound at 2000 Hz. The chromagram of audio signal is shown as:

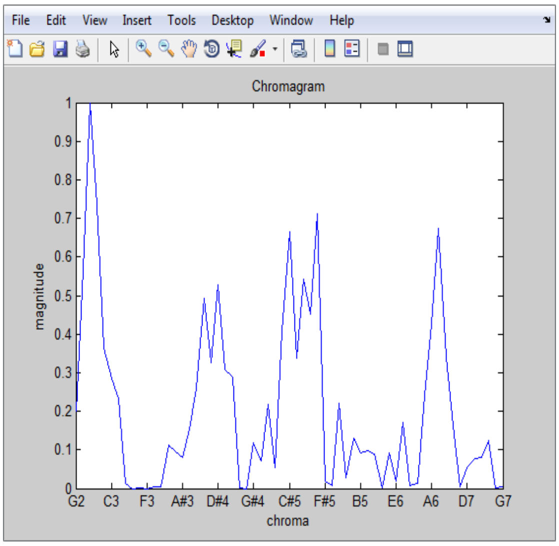

Fig. 1: *The Chromagram Features of Audio.*

Figure 1 presents the chromagram features of audio of 'Asa Raag' from classical Punjabi folk music. The corresponding pitch class is labeled across each chroma feature. One can the amplitude corresponding the each pitch (such as C, C#, D, D#, .. A, B)[8].

CLASSIFICATION USING ARTIFICIAL NEURAL NETWORKS

Artificial Neural Network

Artificial neural Network is family of automatic learning algorithms based on biological nervous system of animals, specifically the brain. ANN is used to estimate the functions which are the expected result for the unknown inputs which are large in numbers [9] [10]. ANN is highly adaptive in nature, thus these are very capable of machine learning along with pattern recognition. To compute the value from input it is represented by system of interconnected "neurons". It is a technique in which machine learns from data provided to it. Network can be trained to the best approximations even if the function is non-linear. The parallel computations are performed which reduces the computation time to the great extent. In this paper, we have used back propagation method.

At the output of n^{th} iteration for j number of neurons, the error signal is defined by

$$e_j(n) = d_j(n) - y_j(n) \quad (4)$$

Corresponding total instantaneous energy Ɛ(n) is given by

$$\mathcal{E}(n) = \frac{1}{2}\sum_{j \in C} e_j^2(n) \quad (5)$$

Where, C is the set which includes all the neurons in output layer of the network. On summing of Ɛ(n) over all n average squared energy is received, and it is normalized with respect to set size N, it is given by,

$$\mathcal{E}_{av} = \frac{1}{N}\sum_{n=1}^{N} \mathcal{E}(n) \quad (6)$$

For any given training set, $\mathcal{E}_{av}$, represents the cost function, which deals with the learning performance. It is needed because it helps to adjust the parameters of the network to have minimum value of $\mathcal{E}_{av}$.

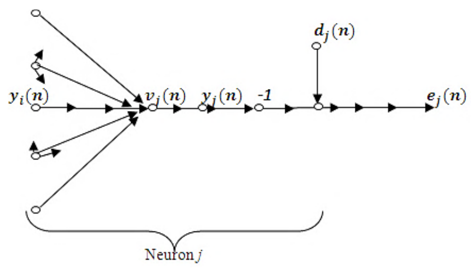

Fig. 2: *Signal Flow Graph Representing the Details of Output j^{th} Neuron.*

Figure 2, represents the neuron j which is fed by a set of signals which are produced by the neuron which are on its left. The induced local field, $v_j(n)$, is produced at the activating function associated with neuron j,

$$v_j(n) = \sum_{i=1}^{m} w_{ji}(n) y_j(n) \quad (7)$$

Where, m is the total number of inputs,

Data Extraction

In this paper, two types of data sets are used for the processing. First set is, the music files are read directly in MATLAB from their WAV file format using wavread. The data extracted has 441,000 values in total for each music sample. Top 256 values are used from each sample. In total 80 samples are used all from different ragas. Second set of music files are taken from the output of chromagram after taking the spectrum of the raw data, using technique explained above. This data is processed data. The processed data is output for each of the sample. The chromagram data has 61 values in each sample.

RESULTS AND DISCUSSION

The data for experiment is collected using various methods. Using first method of data extraction, the extracted data is passed through the neural network algorithm. The output is compared to the desired output. Results are obtained in terms percentage after comparing it with the desired result. Same process is repeated with the data extracted from second process.

Training and testing the neural network.

As described above total of 80 samples were taken for the experiment, from the available dataset 75% i.e. 64 samples were used for training and rest for test. As back propagation method is used for network training and testing, learning rate and moment constants are varied continuously along with variation in number of neurons used. The number of neurons used is varied from 2 to 9. Best result in terms of performance is achieved using 3 neurons. The values of learning rate and moment constant are: 0.1, 0.4, 0.7 and 1 each. Total of 16 combinations are formed out of the moment constant and learning rate. Thus the total performance outcomes for each case of learning rate and moment constant are compared. After checking the performance of training, simulation output is also checked. 25% of dataset is used for the purpose of testing. The percentage classification is done by comparing the simulation output with the required output. A table is formed for simulation output for both the inputs. Thus using the table data box plot is plotted for each case. All the weights are saved and checked for each corresponding value of learning rate and moment constant.

Value of learning rate and moment constant	Simulation output of Raw data (in percentage)	Simulation output of extracted chroma data (in percentage)
LR=0.1, MC=0.1	31.25	18.75
LR=0.1, MC=0.4	31.25	18.75
LR=0.1, MC=0.7	43.75	18.75
LR=0.1, MC=1	56.25	31.25
LR=0.4, MC=0.1	56.25	43.75
LR=0.4, MC=0.4	6.25	12.5
LR=0.4, MC=0.7	18.75	18.75
LR=0.4, MC=1	12.5	43.75
LR=0.7, MC=0.1	31.25	12.5
LR=0.7, MC=0.4	50	6.25
LR=0.7, MC=0.7	6.25	37.5
LR=0.7, MC=1	12.5	43.75
LR=1, MC=0.1	25	68.75
LR=1., MC=0.4	43.75	18.75
LR=1, MC=0.7	6.25	18.75
LR=1, MC=1	12.5	37.5

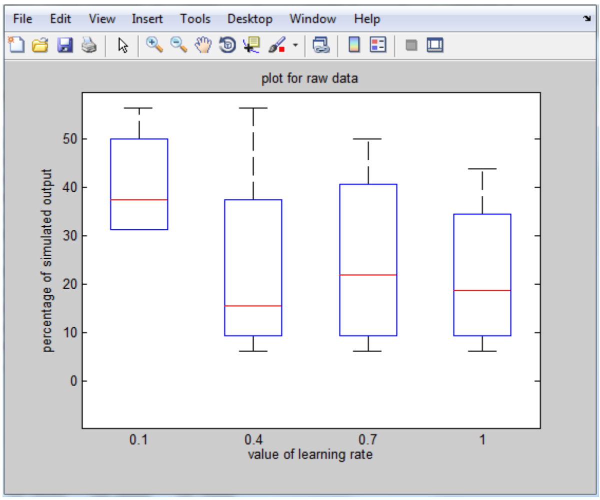

Fig. 3: *Plot for Variation of Simulated Output for Varying Values of Learning Constant in Case of Raw Data.*

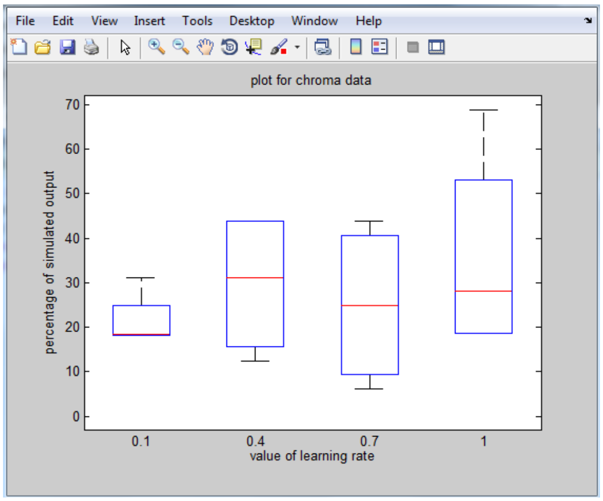

Fig. 4: *Plot for Variation of Simulated Output for Varying Values of Learning Constant in Case of Processed Chroma Data.*

CONCLUSION

The raw data set is not the best data on whose basis we can classify the music signals. These are just like random numbers with no details of their own. The preprocessing done on these random numbers (raw data) give the about the music. Thus can say the chromagram can help to extract the chroma features of the music files. But there is a flaw in this process too. We can undoubtedly extract the chroma features but we can't differentiate exactly which chroma features belong to which of the class of raga. Neural Network's maximum result is 68.75% which doesn't make it the perfect way to classify the classes.

REFERENCES

1. Meinard Müller, Daniel P. W. Ellis, Anssi, Gaël Richard. Signal Processing for Music Analysis. *IEEE Journal of Selected Topics in Signal Processing*. 2011; 5(6): 1088–1110p.
2. Laurent Oudre, Cédric Févotte, Yves Grenier. Probabilistic Template-Based Chord Recognition. *IEEE Transactions on Audio, Speech, and Language Processing*. 2011; 19(8): 2249–.
3. 2259pPasi Saari, Tuomas Eerola, Olivier Lartillot. Generalizability and Simplicity as Criteria in Feature Selection: Application to Mood Classification in Music. *IEEE Transactions on Audio, Speech, and Language Processing*. 2011; 19(6): 1802–1812p.
4. Meinard Müller, Sebastian Ewert. Towards Timbre-Invariant Audio Features for Harmony-Based Music. *IEEE Transactions on Audio, Speech, and Language Processing.* 2010; 18(3): 649–662p.
5. Yonatan Vaizman, Brian McFee. Codebook-Based Audio Feature Representation for Music Information Retrieval. *IEEE/ACM Transactions on Audio, Speech, and Language Processing.* 2014; 22(10): 1483–1493p.
6. Olivier Lartillot, Petri. MIR in MATLAB (ii): a toolbox for musical feature extraction from audio. *Proceedings Conference ISMIR*. 2007: 127–13p.
7. Hideyuki Tachibana, Nobutaka Ono, Hirokazu Kameoka, Shigeki Sagayama. Harmonic/Percussive Sound Separation Based on Anisotropic Smoothness of Spectrograms. *IEEE/ACM Transactions on Audio, Speech, and Language Processing.* 2014; 22(12): 2059–2073p.
8. Pranay Dighe, Parul Agrawal, Harish Karnick, Siddartha Thota ,Bhiksha Raj. Scale Independent Raga Identification Using Chromagram Patterns And Swara Based Features. *IEEE International Conference on Multimedia and Expo Workshops (ICMEW).* 2013:1–4p.
9. Siddharth Sigtia, Simon Dixon. Improved Music Feature Learning with Deep Neural Networks. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).*2014; May: 6959–6963p.
10. Benyamin Matityaho, Miriam Furst. Neural network based model for classification of music type. *Eighteenth Convention of Electrical and Electronics Engineers in Israel.*1995; March: 4.3.4/1–4.3.4/5p.

Cite this Article
Priya, Ravi Kumar. Identification of individual ragas using artificial neural network classifier. *Journal of Artificial Intelligence Research & Advances*. 2015; 2(2): 24–28p.