Genre classification of songs using neural network

Conference Paper · September 2014

DOI: 10.1109/ICCCT.2014.7001506

Sarfaraz Masood
Jamia Millia Islamia

Genre Classification of Songs Using Neural Network

AnshumanGoel, Mohd. Sheezan
JamiaMilliaIslamia
Department of Computer Engg.
Faculty of Engg.& Technology, New Delhi
leocrimson@rediffmail.com
sheezanjmi@yahoo.in

Sarfaraz Masood, AadamSaleem
JamiaMilliaIslamia
Department of Computer Engg.
Faculty of Engg.& Technology, New Delhi
+91 9654774604
smasood@jmi.ac.in
aadam.saleem@gmail.com

ABSTRACT

The objective in this work here is to eliminate the manual work of classifying genres of song in each song. In this work songs can be classified in real-time and proposed parallel architecture can be implemented on the multi-processing system as well. For this purpose, a set of features like beats, tempo, energy, loudness, speechiness, valence, danceability, acousticness, DWT etc. are obtained, using Echonest libraries and are fed into the Parallel Multi-Layer Perceptron Network for the purpose of classification of the genres of the songs. The proposed scheme gave an accuracy of about 85% while classifying the songs for two separate well known genres of Indian Hindi songs, i.e. Sufi and Classical.

Keywords

Genre, Classification, Songs, Echonest, Multi-Layered Perceptron, Artificial Neural Networks, Backpropagation.

1. INTRODUCTION

A large amount of human effort is needed when classifying songs in a music library. People around the world these days, like and prefer to listen to the songs of a particular genre. Many online music commercial sites are providing this feature but this task is done manually till now. The tedious task of first listening to song and then deciding the song to which genre it belongs to doesn't suit the 21st century. To eliminate this we here have tried to propose a new scheme on the genre classification of songs.

2. LITERATURE REVIEW

A recent work in this field of genre classification of songs is done by LindasalwaMuda, MumtajBegam and I.Elamvazuthi, [1] and Carlos N. Silla Jr. and Alessandro L. Koerich [5] which was based upon Mel-frequency Cepstral Coefficients (MFCC). The highest percentage accuracy acquired is around 65%.

The work done by Masato Miyoshi, Satoru Tsuge, TadahiroOyama and Momoyo Ito was based upon the mood of the song [7] i.e., valence was proposed which is an interesting feature that can be extracted from a song.

Benyamin Matityaho, and Miriam Furst proposed a scheme based upon the Fast Fourier Transform (FFT) [2]. This work proposes a method to provide 100% accuracy. Alessandro L. Koerich and Cleverson Poitevin [3] propose to divide the songs into the segments and then extracting feature finds the accuracy in between 73% to 95% for different segments. Zhouyu Fu, Guojun Lu, Kai Ming Ting, Dengsheng Zhang [6] using Naïve Bayes Classifier produces an accuracy of 77%. VikramjitMitra and Chia J. Wang [4] proposes a set of features for classing songs genre with an accuracy of 84% using features beat information, MFCC, DWT, energy, and power. Extending it to using a different Neural Architecture and new sets of features like acousticness and valence, we also reduced the feature vector length with a great extend which makes our system more faster and reliable while working in a distributed environment.

3. PROBLEM DESCRIPTION

The 21st century generation is most adhered to music. But music is not a new age innovation but an old innovation for entertainment and infotainment. Thus, cultural differences around the world give rise to different tastes all around the world. The people around the world like different genres of songs and it depends upon individual too. If we look at the different song websites like saavn, gaana, hungama, etc., offer different genres of songs to there customer. But classifying them is a difficult task for them as it is done manually. It takes human labor in this computer generation. To eliminate this human labor we get a challenge ahead of us to design an automatic genre classifier for songs.

4. FEATURE EXTRACTION

A total of 8 features were used for the classification of songs that are number of beats per minute (bpm), loudness, energy, danceability, speechiness, valence, acouticness and discrete wavelet transform is used.

4.1 Beats Periodicity

The first thing that comes to every music lover is the beats of a particular song. Technically, it has been observed that most of the songs in a particular genre have the same beat periodicity. That is why our first feature is beats per minute. Along with the bpm average, and its variance and skewness are also determined so that entire model of a song can be approximated well.

4.2 Loudness

Loudness is an important factor we look at the songs of different genres. Compare a classical or a jazz song with that of the Rap or Pop. The later are found to have greater loudness then the earlier one. The loudness is determined by combining each segment of the sound then determining the average and variance.

4.3 Energy

Energy of a song is an important factor as if classical song compared with a Rock song will always have lower energy levels. It can be even seen that in some songs the distribution of energy is also even. Taken about this fact, energy is also an important factor which can help in the classification of songs according to their genres.

4.4 Speechiness

Speechiness is defined as the voice content in the song. The songs with high voice content like Rap songs have higher value then compared to other genres. This speechiness factor can also be linked to MFCC coefficients which also gives the information

**Pre-Published Copy*

about the vocal content in the song. MFCC was to develop to represent human ear. MFCC were mainly used in the speech recognition, but it can be used for genre classification of songs because of the speech patterns found in the songs. MFCC is based upon the Fast Fourier Transform (FFT).

4.5 Acousticness
Acousticness differentiates a natural sound with that of the electrical sound. Natural sound of drum sticks, harmonium etc., are found in the Classical or Sufi songs while in the Rock songs electrical guitar sound is more prevalent. Acousticness is calculated from Timbre value of the songs calculated over all the segments of the song.

4.6 Valence
In [7] mood of the song plays a good role in the classification of songs. To make our model more accurate valence i.e., mood of the song is also derived. It determines the positiveness displayed by the song i.e., cheerful, happiness, sad, depressed or angry.

4.7 Danceability
Danceability is defined as the ease with which a person can dance on that particular song. Its very difficult for someone to dance on a sad song as compared to cheerful Rock song. Its much more easier for a person to dance on Rock or Pop song as compared to the Sufi song. It can be calculated as measure of tempo, beat strength or many more parameters.

4.8 Discrete Wavelet Transform (DWT)
DWT is used to recognize various hidden patterns which are not seen by the naked eye. There are many wavelet transforms but here, only two i.e., Haar and Daubechies are used. Like [4], the various coefficients derived are partitioned into six segments. On every partition mean and variance are calculate for both the wavelet transforms. Thus, generating a total of twenty four values that will also be fed into the neural network along with the other above calculated values. It may seem that calculating this feature is not at all good idea. But after having a view of these coefficients of songs of different genres gives a clear view of the importance of the DWT Coefficients.

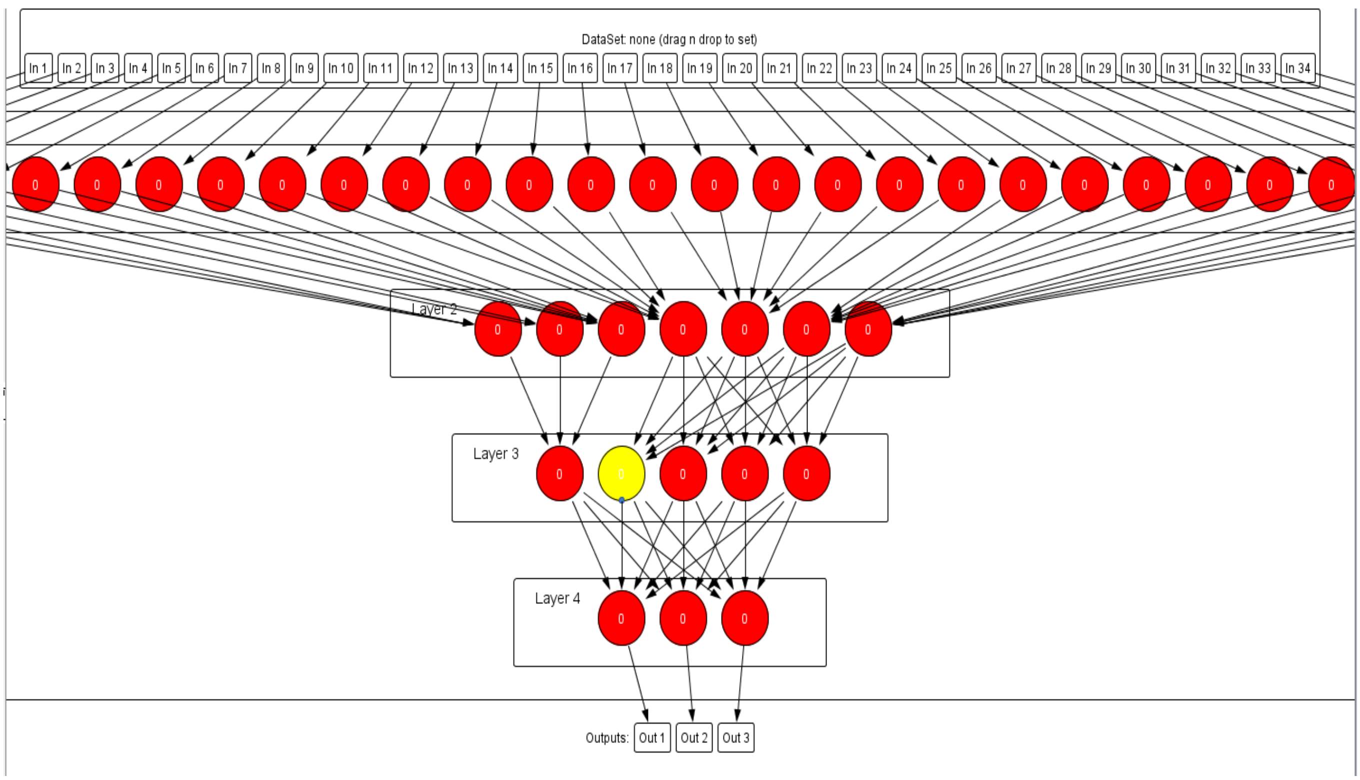

Fig. 1 Neural Network Architecture

5. SYSTEM ARCHITECTURE
The system architecture used here is the Parallel Multi-Layered Perceptron. It is works on Back Propagation Algorithm with momentum. A supervised learning is performed using sigmoidal activation function. The neural network model used is shown in Fig. 1. This model is constructed using Neuroph Studio. Neuroph Studio provides a Netbeans like structure allowing to create dynamic Neural Network links according to the need. It uses JAVA Neuroph API to implement Neural Network functioning. Here there are 34 inputs (three for beats i.e., bpm, beats variance and beats skewness, then loudness mean and average, energy, speechiness, acousticness, valence, danceability, twenty four from DWT Haar and Daubechies mean and variance. With this different architecture every type of feature is evaluated separately according to its area. Like DWT coefficient at extreme right are evaluated in disjunction with other inputs so that they update their weights according to there area thus not affecting the weights of other input areas if there is some robust input even. As supervised learning is adopted the output of this neural network is compared to targeted output and then according to it weights are updated using Back-Propagation Algorithm.

6. RESULTS
The results are evaluated for two genres of songs that are Classical and Sufi songs. The database consists of 200 songs each

of Classical and Sufi. 60% of the songs are used as training set while the rest are used for testing. The database is constructed upon the Indian songs. Echonest Library is used for extraction of some features like acousticness, danceability, valence etc. The echonest library return the URL of a json page which contains the basic information derived from every segment of the songs which are timbre, beats, loudness. From this basic info a number of features are evaluated that are bpm, variance, skewness, loudness etc. Using this, the values are extracted for each song and are fed to the neural architecture created in Neuroph Studio as shown in Fig. 1. The results are evaluated using a learning rate of 0.1 and momentum as 0.5. The network error graph is shown in Fig 2. The graph is drawn for nearly 500 iterations and using training set data. The converges to an error of 0.04 with minimum error of nearly 0.035 around the 250th iteration.

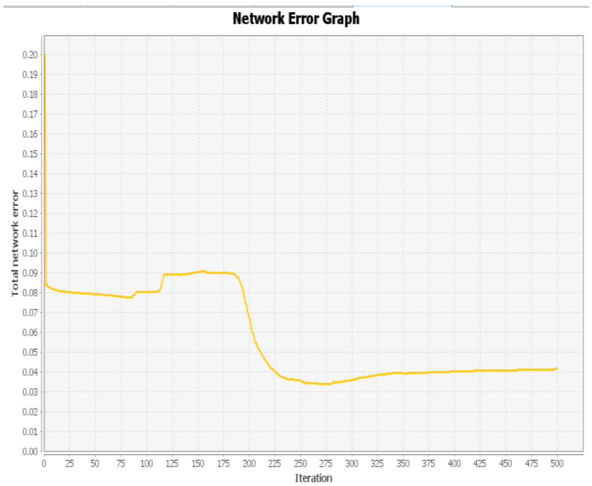

Fig. 2 Network Error Graph

After that the testing is performed using the test data and results are evaluated with an accuracy of 85%. The accuracy of Classical songs is 87% while that of Sufi songs is 82%.

7. CONCLUSION

The propose work uses a total of eight features and applied upon the two genres that are Classical and Sufi songs. The results obtained with an overall accuracy of 85%. The structure designed can be implemented on distributed or multi-processor environment to determine the results faster i.e., on real or run time only. A small in decrease in the accuracy can be because of the fact that Indian Sufi songs are much like Classical songs because of the similar use of natural songs and speechiness patterns. This work can be successfully implemented for other genres around the world like Jazz, Rock, Pop, Rap, etc. It can also be used be used in the music selling websites allowing them to reduce their overhead of classifying and increasing the success rate among their customer because manual work may still have confusion among the operator about the different genres and also require them to have prior knowledge about it. With this no prior knowledge is need. This work can also be extended for songs which are a mixture of two genres like Jazz-Rock, Rock-Rap, Classical-Rock etc.

8. REFERENCES

1. LindasalwaMuda, MumtajBegam and I.Elamvazuthi, "Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques," JOURNAL OF COMPUTING, VOLUME 2, ISSUE 3, MARCH 2010, ISSN 2151-9617.
2. Benyamin Matityaho, and Miriam Furst, "NEURAL NETWORK BASED MODEL FOR CLASSIFICATIONOF MUSIC TYPE."
3. Alessandro L. Koerich and CleversonPoitevin, "Combination of Homogeneous Classifiers for Musical Genre Classification."
4. VikramjitMitra and Chia J. Wang, "A Neural Network based Audio Content Classification,"Proceedings of International Joint Conference on Neural Networks, Orlando, Florida, USA, August 12-17, 2007.
5. Carlos N. Silla Jr. and Alessandro L. Koerich, "Feature Selection in Automatic Music Genre Classification,"Tenth IEEE International Symposium on Multimedia.
6. Zhouyu Fu, Guojun Lu, Kai Ming Ting, Dengsheng Zhang, "Learning Naive Bayes Classifiers for Music Classification and Retrieval," 2010 International Conference on Pattern Recognition.
7. Masato Miyoshi, Satoru Tsuge, TadahiroOyama and Momoyo Ito, "Feature Selection Method for Music Mood Score Detection," 978-1-4577-0005-7/11/$26.00 ©2011 IEEE.

Pre-Published Copy