Music Genre Classification using Improved Artificial Neural Network with Fixed Size Momentum

Nimesh Prabhu
Computer Engineering Department
Goa College of Engineering
Goa, India

Ashvek Asnodkar
Computer Engineering Department
Goa College of Engineering
Goa, India

Rohan Kenkre
Computer Engineering Department
Goa College of Engineering
Goa, India

ABSTRACT

Musical genres are defined as categorical labels that auditors use to characterize pieces of music sample. A musical genre can be characterized by a set of common perceptive parameters. An automatic genre classification would actually be very helpful to replace or complete human genre annotation, which is actually used. Neural networks have found overwhelming success in the area of pattern recognition. The standard back propagation algorithm is used for training network with fixed learning rate. This paper classifies music into genres using improved neural network with fixed size momentum. Finally we validate the proposed algorithm with experimental results of accuracy.

Keywords

Neural network, learning rate, music genre classification, Back Propagation.

1. INTRODUCTION

Browsing and searching by genre can be very effective tools for users of rapidly growing network music archives. The current lack of generally accepted automatic genre classification system necessitates manual classification, which is both time consuming and inconsistent.

Developments in Internet and broadcast technology enable users to enjoy large amounts of multimedia content. With this rapidly increasing amount of data, users require automatic methods to filter, process and store incoming data. A major challenge in this field is the automatic classification of audio.

During the last decade, several authors have proposed algorithms to classify incoming audio data based on different algorithms. Most of these proposed systems combine two processing stages.

Neural networks have found overwhelming success in the area of pattern recognition. Due to the time required to train a Neural Network, many researchers have devoted their efforts to developing speedup techniques [1 – 7]. The neural network can be trained to discern the different criteria's used to classify into classes, and it can do it so in a generalized manner allowing accurate classification of the inputs which are not used during training.

The purpose of this paper is to do feasibility study of a music genre classification system based on music content using an artificial neural network. The second section introduces to related work, third section to framework, fourth section to standard neural network algorithm, fifth section to improved neural network algorithm, sixth section to experimental results and seventh to conclusion and future work.

2. RELATED WORK

The heart of automatic musical classification or analysis system is through the process of extraction of features. Though different classifiers have been compared [8], the choice of features has a large much effect to the recognition accuracy than the selected classifiers have. Even if artificial neural networks classifiers give satisfactory scores many different sets of parameters have been proposed so far. A large number of them are mainly originating from speech recognition or analysis area. There are a wide variety of different features that can be used to characterize audio signals. They are basically time-domain and frequency domain (spectral) features.

Norhamreeza Abdul Hamid and Mohd Najib Mohd Salleh have proposed Improvements in Back Prorogation Algorithm Performance by adaptively changing the gain parameter of the activation function together with Momentum Coefficient and Learning Rate [9]. This hastens up the convergence as well as slide the network through shallow of local minima.

Kavita Burse, Manish Manoria and Vishnu P. S. Kirar have proposed Improved Back Propagation Algorithm to Avoid Local Minima in Multiplicative Neuron Model [10] by the addition of Proportion Factor term helps in convergence of the algorithm five times faster where proportional factor is difference between output and target.

M. T. Fardanesh and Okan K. Ersoy have proposed Classification Accuracy Improvement of Neural Network Classifiers by Using Unlabeled Data [11] by increasing the number of training data; the network makes use of testing data along with training data for learning. It is shown that including the unlabeled samples from underrepresented classes in the training set improves the classification accuracy.

3. FRAMEWORK

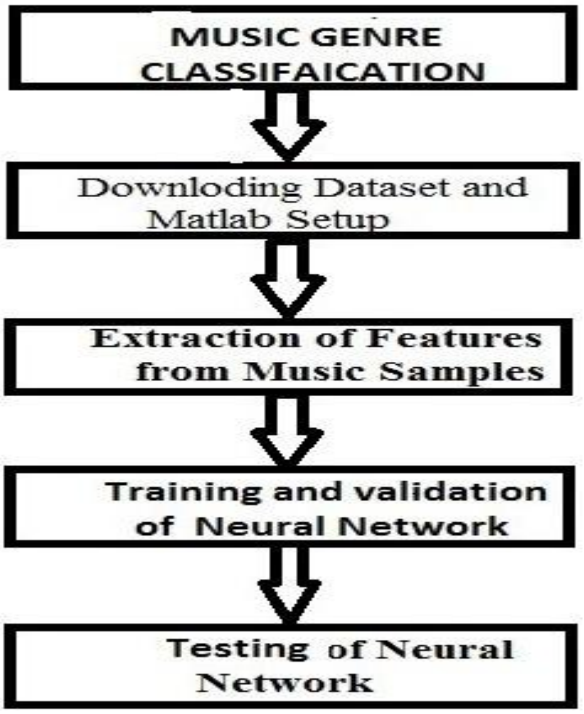

MUSIC GENRE CLASSIFAICATION
Downloding Dataset and Matlab Setup
Extraction of Features from Music Samples
Training and validation of Neural Network
Testing of Neural Network

Figure 1:Design of overall Process

Figure 1 describes framework of whole process. First step is process of downloading dataset and installing matlab. Second step is feature extraction process. Third step is training and validation of neural network both standard and improved. Fourth step is testing of neural network.

4. NEURAL NETWORK ALGORITHMS

The Artificial Neural Network is a near perfect simulation of the biological neural system that is found in humans and other animals. The network is composed of three layers viz. one input, one or more hidden, and a output layer. The number of neurons in the input layer is equal to size of the feature vector. The number of neurons of the output layer is equal to number of classes to be classified. Each neuron in the neural network has a threshold function (activation function) of its own which limits the value of its output. The weights between the neurons and the bias are calculated iteratively. Back Propagation Algorithm is used for training the network. After training the network should be validated and tested.

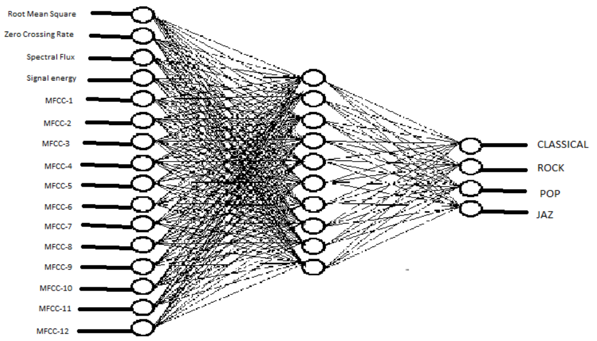


Figure 2: Neural network structure

Figure 2 gives basic idea overall idea of all features which are inputs to input layer, number of neuron in hidden layer and numbers of neurons in output layer with the class in which it classifies.

5. PROPSED METHOLODOGY

Our proposed Neural Network Structure consists of 16 neurons in the input layer which is equal to the number of features which are extracted from the sample dataset. The Output layer consists of 4 neurons so as to classify the dataset into 4 music genre viz. jazz, metal, classical, and pop. The hidden layer consists of 10 neurons which is the average number of neurons in the input and output layer. The weights and bias are randomly initialized n the network. The change in weight and bias are iteratively computed until error is reduced. Out of the total lot of 400 samples of Dataset, 200 samples are used for training using Back Propagation algorithm, 100 samples are used for validation and 100 samples are used for Testing.

The first stage analyzes the incoming waveform and extracts certain parameters (features) from it. The feature extraction process usually involves a large information reduction. The second stage performs a classification based on the extracted features.

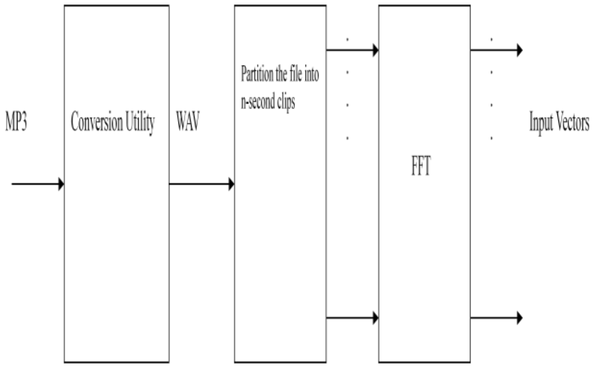


Figure 3: Feature Extraction Process

The above figure 3 describes overall process of feature extraction. First step is conversion utility, second step partition the files n-second clips, third step is applying FFT which gives output which is feature vector which is input to neural network.

5.1 Dataset

First we need a dataset of music files to extract features. Marsyas (Music Analysis, Retrieval, and Synthesis for Audio Signals) is an open source software framework for audio processing with specific emphasis on Music Information Retrieval Applications. GTZAN Genre Collection, of 400 audio tracks each 30 seconds long. There are 4 genres represented, each containing 100 tracks. All the tracks are 22055Hz Mono 16-bit audio files in .wav format. We have chosen four of the most distinct genres for our research: classical, jazz, metal, and pop because multiple previous work has indicated that the success rate declines when the number of classifications is more.

5.2 Feature Extraction

A MP3 file is converted into WAV using wav converter software. A 30 seconds audio file stored in WAV format which is passed to a feature extraction process. The WAV format for audio is simply the right and left stereo signal samples. The feature extraction process calculates 16 numerical features that characterize the particular sample. One of the feature is MFCC that again gives 12 values. Hence, in total 16 values are used to classify the music genres classification(MGC). Feature extraction process is carried out on many different WAV files to create a matrix of containing column's of feature vectors. feature extraction matrix is used to train Neural network.

5.3 Some features that will be extracted.

5.3.1 Zero Crossing Rate:

The Zero crossing rate is the rate of sign-changes along a signal, i.e., the rate at which the signal changes from negative to positive or positive to negative. This feature has been used heavily in both speech recognition and music information retrieval, being a key feature to classify percussive sounds.

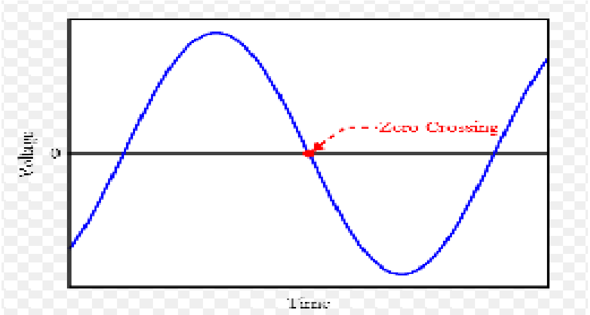


Figure 4: Zero Crossing Rate

$$\mathbf{ZCR} = \frac{\mathbf{1}}{\mathbf{T-1}}\sum_{\mathbf{t=1}}^{\mathbf{T-1}} \mathbf{II}\{\mathbf{s_t s_{t-1}} < 0\}$$

Where S is a signal of length T and the indicator function II{A} is 1 if its argument A is true and 0 otherwise.

5.3.2 Spectral Flux:

Spectral flux is a measure of how quickly the power spectrum of a signal is changing. It is calculated by comparing the power spectrum for the current frame against the power spectrum from the previous frame. It is usually calculated as the 2-norm between the two normalized spectra.

$$\textbf{Spectral flux} = (\boldsymbol{F(n)_t} - \boldsymbol{F(n)_{t-1}})$$

Where $F(n)_t$ and $F(n)_{t-1}$are normalised magnitudes of Fourier transform at current frame t and previous frame t-1.

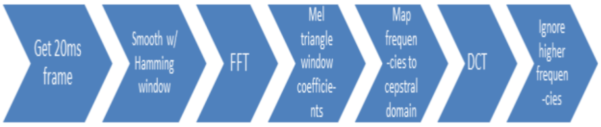
5.3.3 Signal energy:

It is total energy of an audio file calculated by following formula:

$$\textbf{Signal Energy} = \sum_{\boldsymbol{n=1}}^{\boldsymbol{N}} |\boldsymbol{x(n)}|^{\mathbf{2}}$$

where x(n) is feature vector

5.3.4 Mel Frequency Cepstral Coefficients:


Get 20ms frame
Smooth w/ Hamming window
FFT
Mel triangle window coefficie-nts
Map frequen-cies to cepstral domain
DCT
Ignore higher frequen-cies

In music genre classification, the Mel frequency Cepstrum is a representation of the short-term power spectrum of a audio. It is based on a linear cosine transform of a logarithmic power spectrum on a non-linear mel scale of frequency.

Mel-frequency cepstral coefficients (MFCCs) are coefficients that collectively make up an Mel frequency Cepstrum (MFC).

5.3.5 Root Mean Square Level (amplitude):

It is used to calculate root mean square level of amplitude of a audio signal for a continuously varying function or for the series of discrete values.

$$\textbf{RMS} = \sqrt{(\boldsymbol{x_1^2 + x_2^2 +} \;\ldots\ldots \boldsymbol{x_n^2)/n}}$$

Where n = number of samples

5.4 Fixed size Momentum

Fixed size Momentum is basically designed to overcome some of the limitations associated with standard back propagation training. In order to speed up training, many researchers augment each weight update based on the previous weight update. This effectively increases the learning rate [12]. Many algorithms use information from previous weight updates to determine how large an update can be made without diverging [12-14]. This typically uses some form of historical information about a particular weight's gradient.

In this paper we have proposed Fixed size momentum algorithm, which increases speedup over standard momentum. Fixed size momentum is designed to use a fixed width history of recent weight updates for each connection in a neural network. By using this additional information which is stored, Fixed size momentum gives significant speed-up with same or improved accuracy.

Standard weight update rule in back propagation algorithm is

$$\boldsymbol{\Delta w_{ij}(t) = \eta\delta_j x_{ji}}$$

Where

i is the index of source node

j is the index of target node

η is the learning rate

δ_j is the back propagated error term

x is the value of the input into the weight.

This update rule is very slow and time consuming.

Fixed size Momentum uses a fixed size window that captures more information than that is used by standard momentum. By using more memory it is possible to overcome some of the limitations.

Fixed size momentum remembers the most recent n updates to weight and uses that information in the current update for each weight. With standard momentum, the error term from previous update is partially applied to next.. In the worst case, some consecutive samples will have opposite updates. This situation can disrupt the momentum that may have built up and it could take longer to train. Fixed size momentum is able to look at a broader history.

Fixed size Momentum Formula

$$\Delta w_{ij}(t) = \eta\delta_j x_{ji} + f(\eta\delta_j x_{ji}, \Delta w_{ij}(t-1), \Delta w_{ij}(t-2), \ldots, \Delta w_{ij}(t-k))$$

There are k+1 arguments to the function *f*, the first is the current update and the remainder are the k previous updates where k is the window size for the Fixed size momentum algorithm.

The proposed formula helps in convergence faster plus increases classification accuracy with increasing size of window of history which is used to train network.

6. EXPERIMENTAL RESULTS

The accuracy was calculated for various learning rate ranging from 0.1 to 0.5 with standard neural network. The highest accuracy of 83% was recorded when learning rate was 0.2 .

Table 1: Accuracy at different Learning rate

Learning rate	Accuracy
0.1	78%
0.2	83%
0.3	82%
0.4	81%
0.5	82%

Table 2: Confusion matrix when learning rate is 0.2

	Pop	jazz	classical	Metal
Pop	21	4	0	0
Jazz	6	19	0	0
Classical	1	5	19	0
Metal:	0	1	0	24

Table 2 is confusion matrix which shows that it classifies 21 out 25 into pop, 19 out 25 in jazz, 19 out 25 in classical and 24 out 25 into metal.

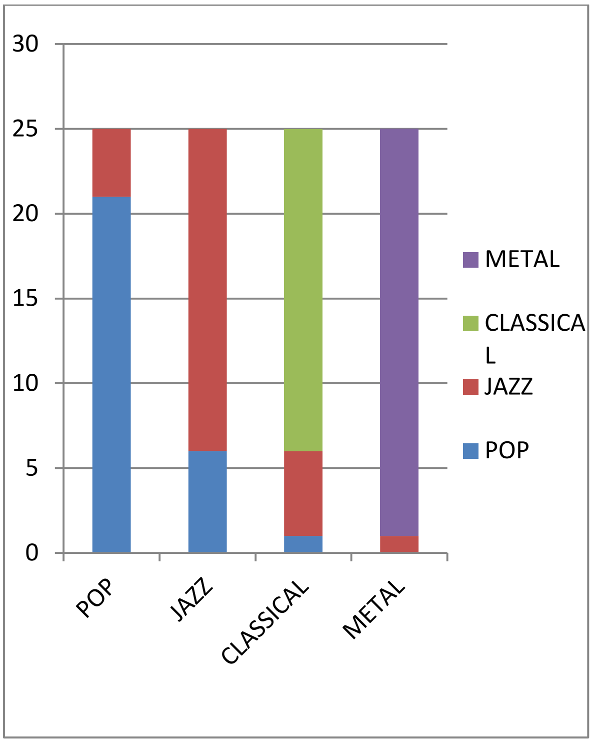


Figure 5: Classification Accuracy of Music into Genres

Figure 5 shows classification accuracy of table 2.

Table 3: Classification Accuracy of Music into Genres by Standard ANN, Standard ANN using momentum, Improved ANN using fixed size momentum.

Accuracy of standard ANN	Accuracy of standard ANN using Momentum	Accuracy of improved ANN using fixed size Momentum
75%	78%	83%
72%	75%	80%
78%	78%	82%
76%	76%	78%
75%	78%	80%

Hence you can observe from Table 3 that you get higher accuracy of classification with improved ANN using fixed size momentum. The above result is obtained using history of size -3 that is it computes weight change using the previous 3 weights and finding average of them and compute new one.

7. CONCLUSION AND FUTURE RESEARCH

The above results state that Jazz and Classical are not classified accurately due to overlapping features in them. It can hence be concluded jazz and classical have more features common in them. Therefore more features have to be extracted to increase more accuracy and classify it more accurately.

Though good results were obtained from the GTZAN dataset, it can be tried for more data sets and extend classification to more genres and even to sub genres . An interesting direction for future research is to associate instrument recognition. Instead of comparing the average from the previous k updates to the current update ,the average can be used in place of the current update. Increasing the fixed size history can be attempted, to increase accuracy.

8. REFERENCES

[1] Leonard, J. and Kramer, M. A.: Improvement of the Backpropagation Algorithm for Training Neural Networks, Computers Chem. Engng., Volume 14, No. 3, pp 337-341, 1990.

[2] Minai, A. A., and Williams, R. D., Acceleration of Back-Propagation Through Learning Rate and Momentum Adaptation, in International Joint Conference on Neural Networks, IEEE, pp 676-679, 1990.

[3] Schiffmann, W., Joost, M., and Werner, R., “Comparison of Optimized Backprop Algorithms”, Artificial Nerual Networks. European Symposium, D-Facto Publications, Brussels, Belgium, 1993.

[4] Silva, Fernando M., & Almeida, Luis B.: “Speeding up Backpropagation”, Advanced Neural Computers, Eckmiller R. (Editor), page 151-158, 1990.

[5] Tollenaere, Tom, “SuperSAB: Fast Adaptive Backpropagation with Good Scaling Properties”, Neural Networks, Vol. 3, pp 561-573, 1990.

[6] Wilamowski, Bogdan W., Chen, Yixin, and Malinowski, Aleksander, “Efficient Algorithm for Training Neural Networks with one Hidden Layer”, Proceedings on the International Conference on Neural Networks, San Diego, CA, 1997.

[7] Jacobs, Robert A., “Increased Rates of Convergence Through Learning Rate Adaption”, Neural Networks, Vol. 1, pp 295-307, 1988.

[8] Norhamreeza Abdul Hamid, Mohd Najib Mohd Salleh(2011).”Improvements of back Propagation Algorithm Performance by Adaptively Changing Gain, Momentum and Learning Rate”In the International journal on New Computer Architecture and Their Applications (UNCAA)1(4):866-878,The Society of Digital Information and Wireless Communications,2011(ISSN:2220-9085)

[9] Kavita Burse, Manish Manoria, Vishnu P. S. Kirar (2010) ” Improved Back Propagation Algorithm to Avoid Local Minima in Multiplicative Neuron Model” In the World Acadamy of Science, Engineering and Technology 48 2010

[10] Jacobs, Robert A., “Increased Rates of Convergence Through Learning Rate Adaption”, Neural Networks, Vol. 1, pp 295-307, 1988.

[11] Minai, A. A., and Williams, R. D., Acceleration of Back-Propagation Through Learning Rate and Momentum Adaptation, in International Joint Conference on Neural Networks, IEEE, pp 676-679, 1990.

[12] Schraudolph, Nicol N., “Fast Second-Order Gradient Descent via O(n) Curvature Matrix-Vector Products”, Neural Computation 2000

[13] Leonard, J. and Kramer, M. A.: Improvement of the Backpropagation Algorithm for Training Neural Networks, Computers Chem. Engng., Volume 14, No. 3, pp 337-341, 1990.

[14] G. Tzanetakis and P.Cook, “Musical Genre Classification of Audio Signals” In IEEE Trans.Acoust. Speech, SignalProcessing , vol.10, ,N°5, July 2002.

[15] Paul Scott, "Music Classification using Neural Networks," Bernard Widrow ,Spring 2001

[16] G. Tzanetakis and P. Cook, “Audio analysis using the discrete wavelet transform” in Proc. Conf. Acoustics andMusic Theory Applications, Sept.2001

[17] H. Murai, M. Okamura, and **S.** Omatu, “Improvement of Pattern Classification Accuracy by Two Kinds of NeuralNetworks", Journal of The Remote Sensing Society of Japan,

[18] Haykin, S. S., "Neural Networks and Learning Machines" New Jersey: Prentice Hall. (2009).

[19] T. Heitolla, "Automatic Classification of music signals ", Master of Science Thesis, February 2003.

[20] R. Duda, P. Hart and D. Stork, “Pattern Classification” , John Wiley & Son, New York, 2000.

[21] M. T. Fardanesh and Okan K. Ersoy" Classification Accuracy Improvement of NeuralNetwork Classifiers by Using Unlabeled Data" in IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, VOL. 36, NO. 3, MAY 1998.