Neural Network Optimization by Genetic Algorithms for the Audio Classification to Speech and Music

Saeed Balochian, Emad Abbasi Seidabad and Saman Zahiri Rad

Department of Electrical engineering, Gonabad branch, Islamic Azad University, Iran
saeed.balochian@gmail.com, emad.abasiseidabad@gmail.com, saman.zahiri.rad@gmail.com

Abstract

In this paper, the execution of some features based on wavelet transform are evaluated through classification of audio to speech and music using the MLP classifiers Optimized by Genetic Algorithm. Classification results show the wavelet features are completely successful in speech/music classification. Experimental comparisons using different wavelets are presented and discussed. By using some wavelet features, extracted from 1-second segments of the signal, we obtained 96.49% accuracy in the audio classification of the MLP classifiers optimized by genetic algorithm.

Keywords: *wavelet, Genetic Algorithm, multi layer perceptron, audio classification*

1. Introduction

Audio segmentation and classification have significant applications in recall of data, archive management, modern human-computer interfaces, and entertainment. One of the basic problems in audio segmentation and classification is speech/music discrimination. The new generation of low bit rate coders and compression technologies may need an estimation of the signal nature in order to achieve a reasonable performance. Therefore, a fast and efficient speech/music discriminator could play a crucial role in such coders to make the signal adaptation task possible and effective.

Researches in closely related areas, such as speech recognition and speaker identification have a long history, while classification and retrieval of audio information are relatively new. Among them is the work Shantha *et al.*, [13] used an improved feature vector formation technique for audio classification and categorization based on wavelet transform. Yang *et al.*, [14], and La Sapienza [1], are used BP neural network to improve the accuracy and intelligence of audio signal's classification and achieved the identification rate of the samples of testing set above 90%.

However, wavelet features have not yet been inquired widely for audio classification. In this paper, we used wavelet based features for audio classification to speech and music by using of the MLP (Multi Layer Perceptron) classifier Optimized by Genetic Algorithm. The new aspect of this work is studying different wavelets with different levels in audio classification tasks. The wavelet features of decomposed levels are tested and evaluated. The wavelet features from 5 levels, including 8 features in each level, are used and compared to conventional features.

One of the most commonly used genetic algorithms optimization method has been proposed, An important property of this algorithm is to find a set of optimal solution of a large collection is no test all possible scenarios.

The paper is structured as follows. Wavelet transform and the feature vectors used in this study are described in Section 2. The classification methods we have employed are presented in Section 3. The system implementation and experimental results are presented in Section 4 and the features and the classification methods are compared in Section 5. The paper is concluded in Section 6.

2. Wavelet Transform and Feature Vectors

In this section, a brief overview of wavelet transform is given and the feature vectors extracted from the wavelet transform are presented.

A. Wavelet Transform

Wavelets are functions that satisfy certain mathematical requirements and are used in representing data or other functions. Wavelet algorithms process data at different scales or resolutions. The wavelet analysis procedure is to adopt a wavelet original model function, called an analyzing wavelet or mother wavelet. Temporal analysis is performed with a contracted, high-frequency version of the original model wavelet, while frequency analysis is performed with a dilated, low-frequency version of the same wavelet [4].

The discrete wavelets transform (DWT) is a new discipline able to giving a time-frequency representation of any given signal. Starting from the original audio signal S, DWT produces two sets of coefficients as shown in Figure 1 [5]. The approximated coefficients A (low frequencies) are produced by passing the signal S through a low pass filter y. The Details coefficients D (high frequencies) are produced by passing the signal S through a low pass filter g.

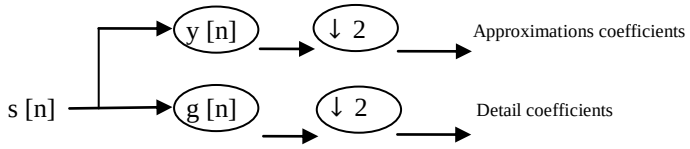

Figure 1. One-level DWT Disintegration [4]

Depending on the application and the length of the signal, the low frequencies part might be further disintegrated into two parts of high and low frequencies. Figure 2 shows a 3-level DWT disintegration of signal S. The original signal S can be reconstructed using the inverse DWT process.

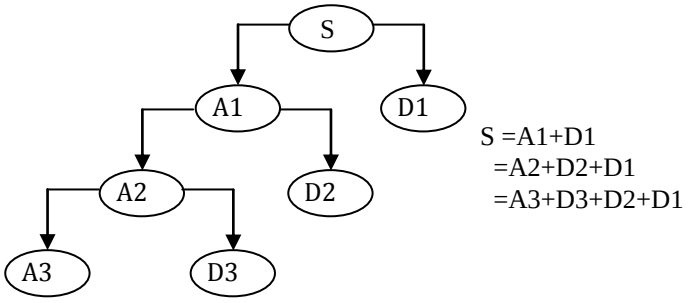

Figure 2. One-level DWT Decomposition [5]

Application Wavelet in the Paper is, Each signal 15 seconds to 15 seconds 1 window to divided and on the each window, wavelet 5 level was used, (According to the above figures), The output value is 6, the details of which 5 and the approximate value is 1. The result, each signal 15 seconds is converted to one vector 15*6. And we are entering the stage of feature vectors.

B. Feature Vectors

In order to obtain a high accuracy in audio classification and segmentation, it is critical to extract features that can capture the major temporal-spectral characteristics of the signals. To classify one-second audio segments, we selected following features in each level of discrete wavelet transform: variance, entropy, standard deviation, maximum, minimum, mean, energy and power.

Spectral analysis shows that pure music is more harmonious than speech, since pure speech contains a respectively of tonal (vowels) and noise-like (consonants) sounds. Speech is characterized by a formantic structure, whereas music is characterized by harmonic structure. The music spectra change slowly, as compared to that of speech. Music can be considered as a succession of periods of relatively stable notes and phones, where speech is rather a mixture of a rapid succession of noisy periods, such as unvoiced consonants, and of periods of relatively stable parts, like vowels. This difference forms our main motivation for applying sinusoidal modeling to the audio classification problem [7].

In this work, the audio signal is divided into non overlapping 1-second segments, and then wavelet decomposition of different wavelets with different levels is performed over each segment of the signal. Therefore the classification algorithm is as the next section.

3. Classification Algorithms

A. Multi Layer Perceptron

A multilayer perceptron (MLP) is a feed forward artificial neural network model that maps sets of input data onto a set of appropriate output. An MLP consists of multiple layers of nodes in a directed graph, with each layer fully connected to the next one. Except for the input nodes, each node is a neuron (or processing element) with a nonlinear activation function. MLP Taking advantage of the supervised learning technique called back propagation for training the network. MLP is a modification of the standard linear perceptron and can distinguish data that is not linearly separable.

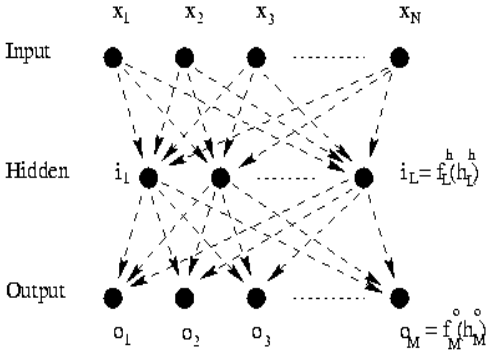

Figure 3. Multi-Layer Perceptron Neural Network [10]

Learning occurs in the perceptron by changing connection weights after each piece of data is processed, based on the amount of error in the output compared to the expected result. This is an example of supervised learning, and is carried out through back propagation, a generalization of the least mean squares algorithm in the linear perceptron [12, 2].

Artificial Neural Networks (ANNs) are used extensively for performing classification tasks due to their reliable and efficient classification performance, parallel computation potential and high adaptability [8]. Based on the philosophy of human nervous system, the ANNs perform extremely well for large and complex feature sets.

Main Result

B. Neural Network Optimized by GA

Genetic algorithms are an iterative process, with iteration of a solution or several solutions work. Figure 4 shows a simple flowchart of the GA.

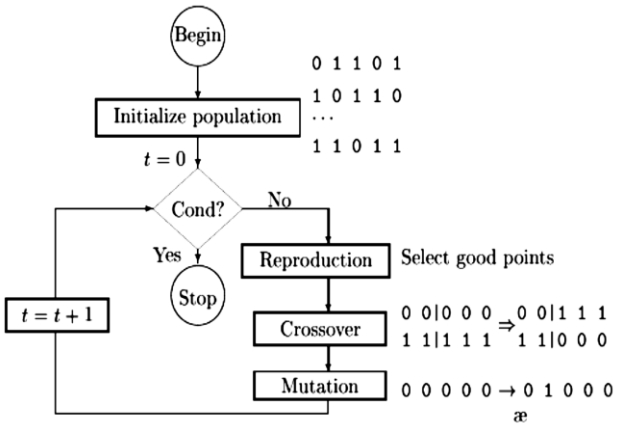

Figure 4. Flowchart of the Genetic Algorithm

GA search starts with an initial random population of solutions. If the criteria final aren't satisfied three different Actuator, Mutation, crossover, selection, are used to update the population. Each iteration of these three Actuator is recognized as of a generation. Because show solutions in Genetic Algorithms much like is a natural chromosome and genetic operators are similar to the GA operators, these names are called genetic algorithms. In fact, the genetic algorithm to solve simple three-step iteration can see in the figure below, the search will be.

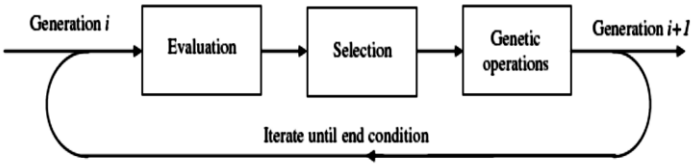

Figure 5. Genetic Algorithm Steps

The first step, called the group of points searched the population 1st is called according to assess the objective function. In the second step, based on the assessment of the condition, in some parts of problem solved are chosen as candidates. In the third step, two genetic operators are applied to these candidates to build the next generation population. The process to obtain the final criterion will be repeated. The final criterion is that an acceptable result or is achieved or the maximum number of generations is repeated.

Neural network based on gradient methods to find the weight that makes it into are the reciprocating.

In the paper, Wei Shen, Mian Xing, are used genetic algorithms in the neural networks used for implementation on a hybrid signal [6].

In this paper, a neural network optimized by genetic algorithms has been discussed. Genetic algorithms for determining the weights and biases of the neural network have been used, which is dramatically less the number of iterations and even versus of changes to no enlarge the network and the network is favorable. This causes the stability of the network and in the applications become more is similar examples.

Algorithm

Step 1: Feature extraction with wavelet:

Wavelet 5 levels with 8 features in each level, which is generally considered 48 features with 1 second windows. Therefore, in this paper the signal xtrain=54680*48, which that desired output matrix with xtrain=54680*48 having ytrain=54680*1.

Step 2:

Creation a five-layer an artificial neural network in the first layer 5 neurons and 15 neurons in the second layer and the last layer 1 neurons are located (15*5*1). The network describes, have a number of the weight that they initially unknown. That the neural network with gradient methods reciprocating to finds the weights, that these methods are very much depends on the size of the network. We've replaced this algorithm with genetic algorithm.

Step 3: Constitute the initial population in genetic algorithms:

The population of a basis weight neural networks that are randomly created. (*i.e.*: per npop, population: npop*351).

Step 4: Compute the target function in algorithm GA:

By putting the population created in the previous step as weights neural network, the output value calculates and with the desired value to compare, *i.e.* Xtrain created in Step 1 and obtain the output of the npop, *i.e.* obtain the $y_o$train to all of the npop calculated. Now, By comparing the $y_o$train(54680*npop) with the target function Primary ytrain, Creation target function GA:

$$e=sum\ (repmat\ (ytrain;npop) - y_o train)$$

Step 5: sort the Descending value error (e).

Step 6: Select the first value in Step 5 to find the Best Answer.

Step 7: If the error was zero, therefore out of the algorithms. And otherwise, the following steps will be.

Step 8: Create the crossover and mutation and new population.

Step 9: back to step 4.

Application of GA in Neural Networks

Neural network to get his weight gradient method uses, and methods in the networks adult is so speed of reduce. In this paper, a GA algorithm is obtain the weights of neural networks, and the reason, a speed increase is significant. The GA objective function is defined as follows:

y_o : The output obtained , y_d : The desired output

The objective function: $Min \sum y_o - y_d$

4. Experiments

The "music-speech" corpus used in this study is a collection of 240 15-sec sound files, randomly selected from

The radio programs [15]. This corpus is taken as a standard benchmark for audio system evaluations and has been used in many audio classification studies [15, 11].

For the feature extraction, the audio signal is partitioned into 1-second segments. Each classifier is evaluated using labeled data sets, each 20 minutes of speech and music data. Each model is trained with 60 15-sec long training speech files (900 seconds) and 60 15-sec training vocal and non-vocal music files (900 seconds). Each system tested over 20 15-sec speech files (300 seconds), 20 15-sec vocal music files (300 seconds), and 21 15-sec non-vocal music files (315 seconds). Thus, each system is trained over 120 15-sec files, (1800 seconds) and is tested with 61 15-sec files (915 seconds). For each 1-second audio segments, we used 8 features in each level of discrete wavelet transform including: variance, entropy, standard deviation, maximum, minimum, mean, energy and power.

The result of audio type classification is a set of 1-second segments, labeled as 'S' or 'M' for speech or music, respectively. A higher overall classification accuracy is achieved by applying a smoothing rule to the resulting frames. Based on this rule, SMS is changed to SSS and MSM to MMM.

The general function which is used for classifying test data was Trainlm. Figures 6 to 9 show the curves, result of training, regression of training, regression of test and result of test in MLP experiments.

The two Figures 6 and 7 are the results of the training data, the two Figure 8 and 9 are the results of the test data. The bold lines in the Figures 6 and 9 are the desired results and the light lines are the obtained results in this paper, and therefore, the Figure 6 shows the good fit to the results of the training data and the Figure 9 shows the number of error for testing data.

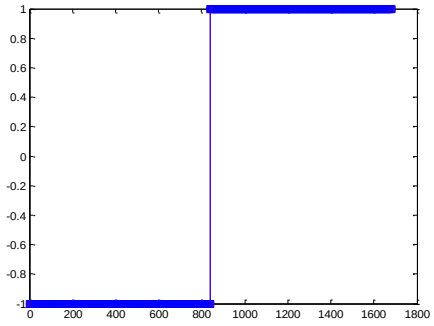

Figure 6. Result of Training

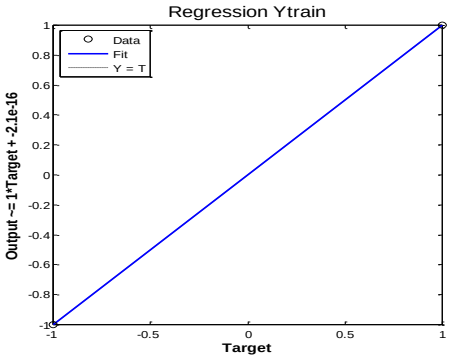

Figure 7. Regression of Training

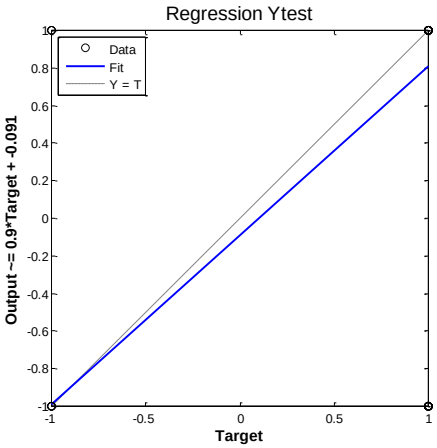

Figure 8. Regression of Test

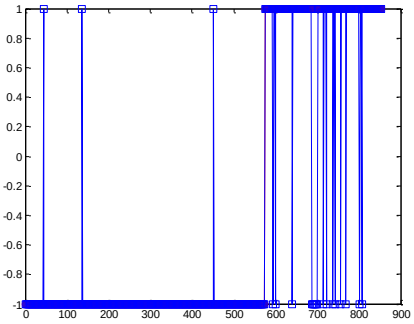

Figure 9. Result of Test

And at following table you can see the total results:

Table 1: The Total Results

	number of the File	Size of the Matrix after Feature Vectors	Number of the False	Percent Error
Train	120 (files 15 sec)	48 * 1680	0	0 %
Test	61 (files 15 sec)	48 * 854	30	3.51 %

5. Results and Discussion

We examined the wavelet decomposition with different levels, and found that the optimal order was 5 when the eight described features were used in each level. Thus, for each 1-second segment, a set of 48 features are generated. We tested different wavelets and found Bior 5.5 wavelet was the best

As a comparison between the used classification methods, neural network optimized by GA is better than the MLP classifier.

Neural network been used includes the 3 layers which the first layer the 5 Neuron and in the layers second and third respectively 15 and 1 Neuron has existed. The first and

second layers are TANSING, and the linear output function as PURLINE. The overall function for classification test data is into TRAINLM that has been used and genetic algorithms have been used for it is 100 iteration and the population of iteration is 150 and on this iteration are 30% mutation and 60% crossover.

6. Conclusions

In this paper, we have studied wavelet transform based features for the audio classification into speech and music. The **execution** of different wavelets and different disintegration levels has been evaluated using both the neural network optimized by GA and the MLP classifiers. Among wavelets, the Bior 5.5 yields a higher performance in the sense of classification error.

References

[1] L. Sapienza, "Audio signal processing by neural networks", Journal ,ELSEVIER, Neurocomputing, vol. 55, **(2003)**, pp. 593-625.
[2] J. Park, F. Diehl, M. J. F. Gales, M. Tomalin and P. C. Woodland, "The efficient incorporation of MLP features into automatic speech recognition systems", Journal, ELSEVIER, Computer Speech and Language, vol. 25, **(2011)**, pp. 519-534.
[3] C. Cortes and V. Vapnik, "Support vector networks," Mach. Learn, vol. 20, **(1995)**, pp. 273-297.
[4] A. Graps, "An Introduction to Wavelets", Journals & Magazines, IEEE Computational Science and Engineering, vol. 2, no. 2, **(1995)**, pp. 50-61.
[5] A. Al-Haj and A. Mohamad, "An Introduction to Wavelet Digital Audio Watermarking Based on the Discrete Wavelets Transform and Singular Value Decompositions," European Journal of Scientific Research, vol. 39, no. 1, **(2010)**, pp. 6-21.
[6] W. Shen and M. Xing, "Signal hybrid HMM-GA-MLP classifier for continuous EMG classification purpose", IEEE, **(2009)**.
[7] S. Ramamohan and S. Dandapat, "Sinusoidal model-based analysis and classification of stressed speech", IEEE Trans. On Audio, Speech and Language Processing, vol. 14, no. 3, pp. 737-746, **(2006)** May.
[8] J. C. Principe, N. R. Euliano and W. C. Lefebvre, "Neural and Adaptive Systems: Fundamentals through Simulations", John Wiley & Sons, Inc., **(2000)** February 29.
[9] V. Vapnik, "The Nature of Statistical Learning Theory", Springer Publications, Book, Springer, **(1999)** November 19.
[10] D. Hoffman, D. Blei, and P. Cook, "Easy as CBA: A simple probabilistic model for tagging music", ISMIR - International Symposium / Conference on Music Information Retrieval Publications: 774 Citation Count, (Self-Citation: 1,688), vol. 10, **(2011)**, pp. 633.
[11] J. Shirazi and S. Ghaemmaghami, "Improvement to speech-music discrimination using sinusoidal model based features", Springer, Multimedia tools and application journal, vol. 50, **(2010)**, pp. 415-435.
[12] T. Marwala, "Finite Element ModelUpdating Using Computational Intelligence Techniques", Book, Springer, **(2010)**.
[13] R. Shantha, D. Sugumar and V .Sadasivam, "Audio Signal Classification Based on Optimal Wavelet and Support Vector Machine", IEEE International Conference on Computational Intelligence and Multimedia Applications, PI. 0-7695-3050-8, vol. 2, **(2007)**, pp. 544-548.
[14] L. Yang and Z. Yang, "Study on Audio Signal's Classification Based on BP Neural Network", IEEE Conference Publications on Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC), **(2011)**, pp. 5153-5155,
[15] E. Scheirer and M. Slaney, "Construction and evaluation of a robust multifeature speech/music discriminator", Proc. ICASSP- 97, **(1997)** April, pp. 21-24.
[16] J. Saunders, "Real-time discrimination of broadcast speech/music", Proc ICASSP-96, **(1996)** May, pp. 993-996.