IJBPAS, May, 2015, 4(5), Special Issue: 284-292

ISSN: 2277–4998

International Journal of Biology, Pharmacy and Allied Sciences (IJBPAS)

'A Bridge Between Laboratory and Reader'

www.ijbpas.com

CLASSIFICATION OF VOICE DISORDERS BASED ON FEATURE SELECTION METHODS USING SUPPORT VECTOR MACHINE (SVM) AND ARTIFICIAL NEURAL NETWORK (ANN)

SARA SOLEIMANPOUR *, REZA SHAHGHADAMI

Department of Biomedical Engineering, Faculty of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran, (Email: Sara.Soleimanpour@gmail.com)

ABSTRACT

Nowadays digital speech processing is one of the most common methods to analyze the vocal tract .Experts analyze human voice for so many purposes such as: diagnosis of larynx disorders, diagnosis of gender, diagnosis of mental states (feeling sad,happy,etc.), welfare and security services and etc , but all of these purposes without digital processing, ends to failure. Scientists use digital speech processing to achieve their goals. In this paper, speech signals recorded from different people are given to the two most important classifiers: SVM and ANN, and then,by using feature selection methods, we reduce the amount of calculations and achieve optimum feature set and at last we compare the performance of the two classifiers. In classifying human voice,paying attention to features which are used in this work , for support vector machine we have gained an average correct classification rate of 70% and for artificial neural network we have gained an average correct classification rate of 50%.

Key words: Signal processing,Voice disorders, Feature selection, Genetic algorithm, Classifier, Artifacial neural networks, Support vector machine.

INTRODUCTION

Phonating occurs in this way: during breathing, exhale air removed from the lungs passes through the trachea, glottis, pharynx and mouth and hits the vocal organs such as vocal cord , temporarily closed vocal tract organs , and different sounds of speech are produced by these changes. In production mechanism, it is possible that 22 muscles change several times during production of a simple sentence. During the speech, hundreds of actions take place per second in order to muscle adjustment .When the wrong muscles contract or when suitable muscles contract more poorly or even more intensely than usual,we will see abnormal control and production speech patterns [1]. The sound which we hear is the conversion of air pulses into physical changes of vocal tract . Some of the speech organs can be seen in figure 1.

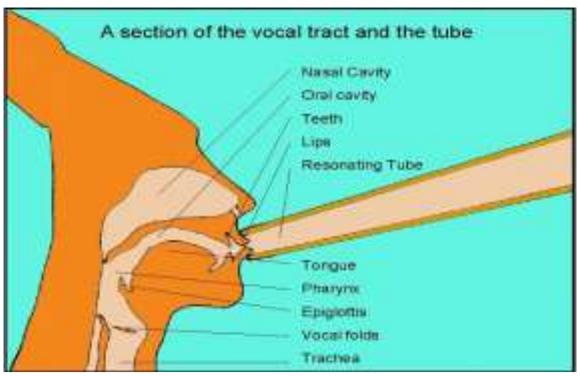


Figure 1 Schematic view of some of speech organs

Effective use of the breathing, phonation, resonance and producing organs is necessary to help to provide the requirements for normal and quick speech .Various cortical and subcortical centers are responsible for coordinating the many muscles involved in speech.

If there is a problem in any or all of the components of speech production system, speaking act will face difficulty. Today, digital speech processing is one of the most common methods for analyzing vocal tract. Diseases associated with the larynx have been mysterious in most cases and a person will not go to the doctor until there is very special changes in his voice,and then he shoud pay so much money for his disease to be diagnosed and cured .

Therefore, having a low-cost and convenient system for people today it is necessary. In this study, we selected and used a combination of the most popular features which can favorably show characteristics of human voice and vocal tract, and we also used the two most popular and common classifiers which have high accuracy in separating different classes . Also, due to high amount of calculations, we used an algorithm called genetic algorithm to reduce the size of feature matrix and reduce the amount of calculations and also to save time.

Research Background

Expanding doing activities in this field returns to the eighties of the nineteenth century. In these years many methods with different functions have been used to detect diseases related to larynx , but the reason that makes researchers to continue using these methods is the accuracy, speed and low amount of calculations of these methods. The ways in which researchers study in this field historically and in terms of applications can be divided into two groups:In early years of studying in this field, the researchers used to find new ways to process the human voice, but later, researchers focused on finding suitable solutions for optimization of those methods . Today, accuracy, speed, easier calculation methods, low amount of calculations, and using low-cost methods are the fundamental factors to which scientists pay attention in the field of human voice processing.

Carlos Hernandez and colleagues [2] used the acoustic parameters as inputs to the neural network and gained the accuracy100% .The neural network techniques have been used to reduce the number of original set of inputs .In this paper, features that have been extracted from MDVP program have been used and researchers have concluded that only two acoustic parameters are sufficient to separate healthy voice from disordered voice.

AlirezaA.Dibazar and colleagues [3] used two databases .The first database contains 710 acoustic speech sample files that belong to patients with organic , neuralgic,traumatic and psychological voice disorders and 53 acoustic speech sample files that belong to healthy people , and all the subjects in this database have uttered the sustained vowel / a /. The second database contains657 acoustic speech sample files belonging to patients with the same diseases as the first database and 53 acoustic speech sample files belonging to healthy people who have red of "Rainbow passage" . Two different methods with different characteristics have been compared with each other in this study. The best results have been obtained with the use of MFCCs and pitch as features and the hidden Markov model as classifier and uttering sustained vowel / a /.

Jagadish Nayak and colleagues [4] used a database containing 50 acoustic speech sample files including 25 files related to healthy people and 25 files related to people with vocal cord paralyses disease .The wavelet analysis have been used to identify voice disorders .The number of test data is 20 including 10 healthy samples and 10 disordered ones .The correct classification rate gained in this work is 100% for normal speech and 80% for pathological speech.

A.Schunk Jr. and colleagues [5] a database containing 64 acoustic speech sample files including 13 files related to healthy people and 51 files related to people with disordered voice and all of the people whom voices are included in the database , have uttered sustained vowel /a/. Sampling frequency is 25 kHz. The wavelet packet transform have been used for feature extraction and we can see that the best level of wavelet decomposition is level 5. Finally the researchers gained the correct classification rate of 84.3%.

Everthon S. Fonsecaand and colleagues [6] have used a database that contains 60 acoustic speech sample files belonging to those who uttered sustained vowel/ a /. 30 acoustic speech sample files belong to healthy people and 30 other belong to disordered voice people .Finally they have reached the accuracy of 95%.

Wenxi Chen and colleagues [7] have suggested a support vector machine based on the classification methods for disordered voice detection .The results show that the classifier is not sensitive to the parameters of the kernel function.

METHODOLOGY

In this paper, we classify human voices with the use of support vector machine and artificial neural network classifiers and at last, we compare the performance of these two classifiers. The important point is the combination of features used in this study which is almost unique compared to similar studies.

The purpose of feature extraction is finding the characteristics of signal which describe it in a more appropriate and suitable way. In this work, we have used a new combination of features and we have also used SVM and ANN classifiers to classify human voice. At last, we have compared the performance of these two classifiers. Features used in this work are as follows: Pitch, Format, Cepstrum coefficients, MFCCs, Two-dimensional cepstrum coefficients, Energy and Entropy extracted from wavelet coefficients [8].

Support Vector Machine

Support vector machine is a learning algorithm which is headed by Vladimir vepenik and his colleagues in the mid-nineties. Support vector machines can be considered as machine learning algorithm families. A large subset of these algorithms is kernel based learning method. This subset can be divided into two methods which are: supervised and unsupervised methods. Support vector machine can be considered as supervised learning method.

The main idea of SVM is that this classifier uses a non-linear function to transfer data from input space to feature space in order to separate classes linearly. Then SVM finds the best separator border, line or page automatically [9] .

To train support vector machine we have used Polynomial (Gaussian) and RBF kernels that are very common in this field.

Artificial Neural Network

Artificial neural networks are simple models that simulate nervous system and have so many applications in solving various scientific problems .ANN is a practical way to learn the various functions such as real amount functions, functions with discrete values and vector functions and is also a way to calculate that can be built by connecting multiple processing units.

Since the artificial neural network is inspired by the human brain, we can say that in fact, nodes simulate the function of nerve cells and the connections between nodes are designed to simulate synapses[10] .

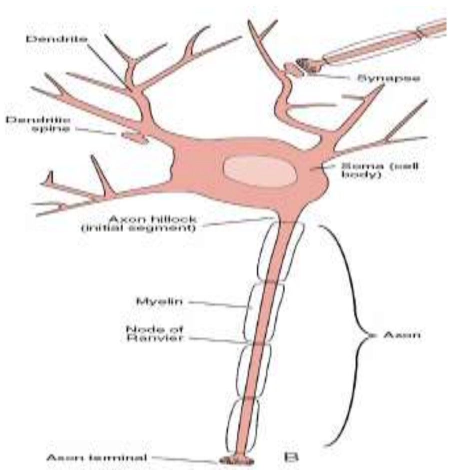


Figure2: A sample of actual nerve cells

To train artificial neural network we have used Levenberg-Marquardt method.

In this study, we have used 155 audio files belonging to the three groups of people .64 files belong to people who suffer from vocal fold paralysis disease , 38 files are related to people who suffer from edema disease and other 53 files are related to healthy people. We have used 70% of data (109 audio files) for training the classifiers and 30% of data (46 audio files) for testing them . To evaluate the performance of the classifiers,we have used CCR (Correct Classification Rate) criterion that can be measured as follow :

CCR = [Sum (Class-estimated = Class-test) / num-test] * 100

From above formula we can see that the correct classification rate for a classifier equals to percentage of ratio of total number of test data whose class is estimated truly, divided by the total number of the test data . In this study we used genetic algorithm to reduce the time needed and computation volume . Total number of features we extracted from the audio windowed signal is 2807 features that we applied them to classifiers in five stages.

RESULTS AND DISCUSSION

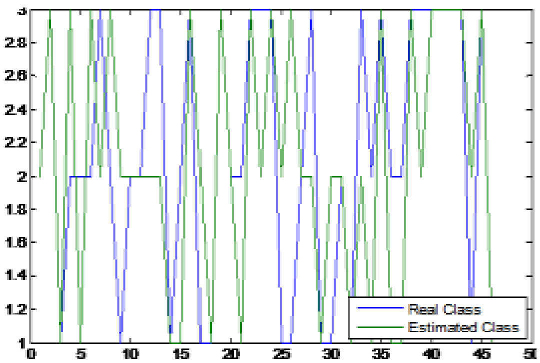
In the findings presented below, CCR represents the rate of correct classification of support vector machine before applying genetic algorithm and feature selection to SVM, CCR feat equals to the rate of correct classification of support vector machine after applying GA and feature selection to SVM and CCR_NN equals to the rate of correct classification of ANN after applying feature selection to ANN. Results for ANN and SVM with RBF kernel with a sigma equal to 6 are as follows:

Table (1) Results for ANN and SVM with RBF kernel

Sum-x	CCR	CCR-feat	CCR-NN

100	32. 6087	86. 9565	63. 0435
200	32. 6087	72. 0496	58. 6957
300	32. 6087	65. 0152	54. 3478
400	32. 6087	57. 5515	50
500	32. 6087	45. 0176	43. 4783

Figures (3) and (4) show examples of the comparison between real class and estimated class for SVM(with RBF kernel) and ANN .



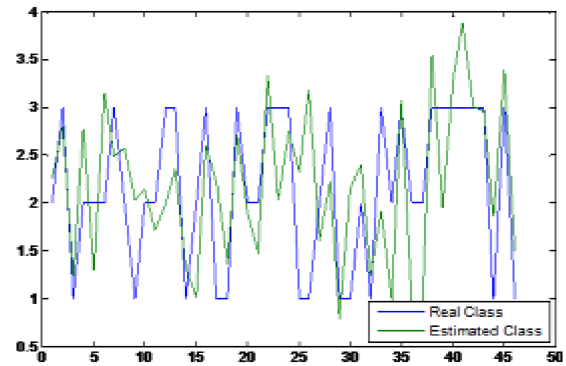


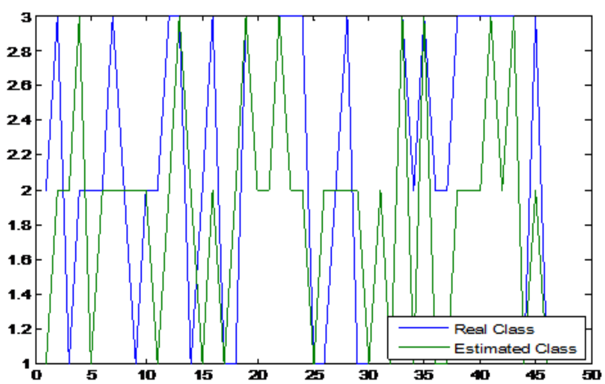
Figure 3 Comparison for SVM for Sum-x=100 **Figure 4 Comparison for ANN for Sum-x=100**

Results for ANN and SVM with Polynominal kernel of degree 3 are as follows:

Table (2) Results for ANN and SVM with Polynominal kernel

Sum-x	CCR	CCR-feat	CCR-NN
100	63. 0435	83. 8509	55. 5900
200	63. 0435	81. 6770	52. 1739
300	63. 0435	80. 4348	47. 5155
400	63. 0435	79. 1925	45. 9627
500	63. 0435	77. 9503	43. 4783

Figures (5) and (6) show examples of the comparison between real class and estimated class For SVM(with polynomial kernel) and ANN.



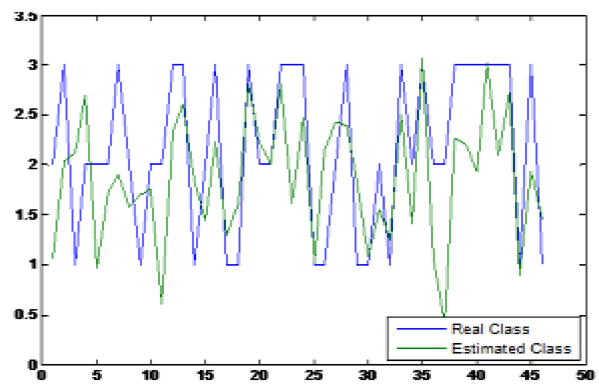


Figure5 Comparison for SVM for Sum-x=100 **Figure 6 Comparison for ANN for Sum-x=100**

The results for SVM with RBF kernel and ANN after applying the optimized features are shown in the following diagram:

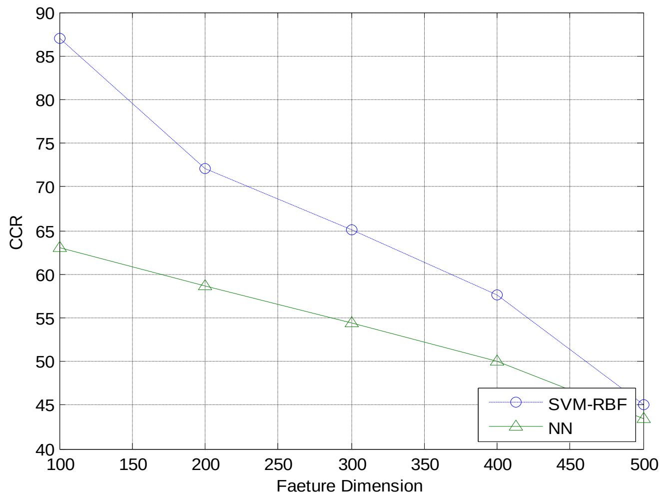


Figure7 comparison between performance of ANN and SVM with RBF kernel

The results for SVM with Polynomial kernel and ANN after applying optimized features are shown in the following diagram:

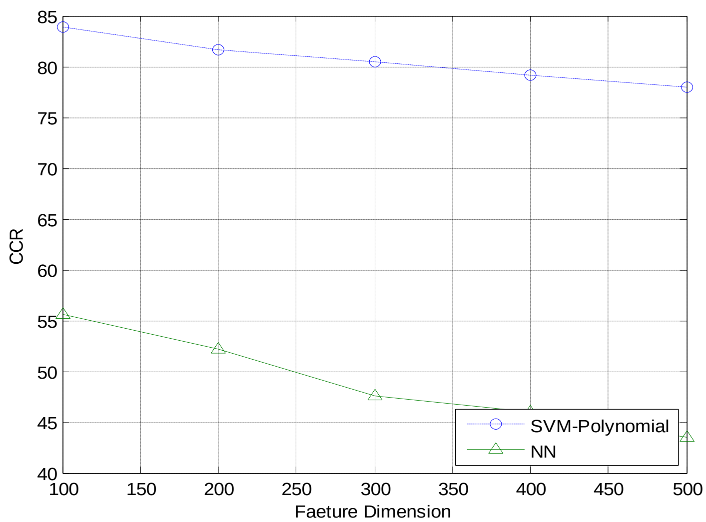


Figure 8: comparison between performance of ANN and SVM with Polynomial kernel

We can see that whenever the number of data is lower, the performance of two classifiers is better and they classify the human voices more accurately but when the number of data increases, the accuracy of classifiers in separating the classes reduces.

In both diagrams we can see that performance of SVM is better than ANN.

The reason of this, is that SVM classifier is basically a two-class classifier and in issues that we have to deal with more than two classes, SVM considers the classes one by one, and investigates whether the data belongs to one of these two classes or not, and it never considers all of the classes together,so we can see that calculations are more complicated in ANN rather than SVM. In the problem we face in this project, we have three classes and therefore, support vector machine must check threeclasses.

REFERENCES

1. Marvin L.editors. Production Speech,4th ed 1998 .

2. Carlos Hernandez-Espinosa,'Mercedes Fernindez-Redondo, Pedro Gomez-Vilda,JuanI.Godi Llorente, Santiago Aguilera-Navarro"Diagnosis of Vocal and Voice Disorders by the SpeechSignal."2000 IEEE.

3. Alireza A. Dibazar, S. Narayanan, T. W. Berger .Feature Analysis for AutomaticDetection of PathologicalSpeech .Processing of the second Joint EMBS/BMES Conference Houston, TX , USA . 23-26 October 2002.

4. JagadishNayak," Identification of Voice Disorders using Speech Samples".2003 lEEE.

5. A. Schuck Jr.’, L. V.Guimarges’and J. 0.Wisbe:ck’."Dysphonic Voice Classification using Wavelet PacketTransform and Artificial Neural Network",2003 IEEE.

6. Everthon S. Fonseca,Rodrigo C.Guido, André C. Silvestre and José Carlos Pereira.Discrete Wavelet Transform and Support Vector Machine Applied to Pathological Voice Signals Identification.IEEE 2005.

7. Wenxi Chen, Cepeng, Xin Zhu, Baikun Wan, Daming Wei.SVM-Based Identification of Pathological Voices. Proceeding of the 29 the Annual International Conference of the IEEE EMBS2007 August;23-26.

8. Lotfi Salhi , Talbi Mourad , Adnene Cherif . Voice Disorders ClassificationUsing Multilayar Neural Network . International Conference on Signals, CircuitsAnd Systems IEEE 2008 .

9. Meisam Khalil Arjmandi, Mohammad Pooyan, Hojat Mohammadnejad, MansourVali. Voice Disorder Identification Based On Different Feature Reduction Methodologies and Support Vector Machine.Proceeding of ICEE 2010 May 11-13.

10. Mahmoud I.Abdalla, Haitham A. Abobakr, Tamer S. Gaafar. DWT and MFCCs Based Feature Extraction Methods For Isolated Word Recognition. International Journal of Computer Application 2013 May; 20(69) .