Gender Classification Based on FeedForward Backpropagation Neural Network

S. Mostafa Rahimi Azghadi[1], M. Reza Bonyadi[1] and Hamed Shahhosseini[2]
1 Department of Electrical and Computer Engineering, Shahid Beheshti University, Evin, Tehran, Iran.
{M_rahimi, M_bonyadi}@std.sbu.ac.ir
2 Department of Electrical and Computer Engineering, Shahid Beheshti University, Evin, Tehran, Iran.
H_shahhosseini@sbu.ac.ir

Abstract. Gender classification based on speech signal is an important task in variant fields such as content-based multimedia. In this paper we propose a novel and efficient method for gender classification based on neural network. In our work pitch feature of voice is used for classification between males and females. Our method is based on an MLP neural network. About 96 % of classification accuracy is obtained for 1 second speech segments.

Keywords. Gender classifications, Backpropagation neural network, pitch features, Fast Fourier Transform.

1 Introduction

Automatically detecting the gender of a speaker has several potential applications. In the content of automatic speech recognition, gender dependent models are more accurate than gender independent ones [1]. Also, gender dependent speech coders are more accurate than gender independent ones [2]. Therefore, automatic gender classification can be important tool in multimedia signal analysis systems.
The proposed technique assumes a constraint on the speech segment lengths, such as other existing techniques. Konig and Morgan (1992) extracted 12 Linear Prediction coding Coefficients (LPC) and the energy feature every 500 ms and used a Multi-Layer Perceptron as a classifier for gender detection [3]. Vergin and Farhat (1996) used the first two formants estimated from vowels to classify gender based on a 7 seconds sentences reporting 85% of classification accuracy on the Air Travel Information System (ATIS) corpus (Hemphill Charles et al., 1990) containing

specifically recorded clean speech[4]. Parris and Carey (1996) combined pitch and HMM for gender identification reporting results of 97.3% [5]. Their experiments have been carried out on sentences of 5 seconds from the OGI database. Some studies on the behavior of specific speech units, such as phonemes, for each gender were carried out [6].

This overview of the existing techniques for gender identification shows that the reported accuracies are generally based on sentences from 3 to 7 seconds obtained manually. In our work, speech segments have 1 second length and we obtained 96 % accuracy.

In several studies, some preprocessing of speech is also done, such as silence removal or phoneme recognition.

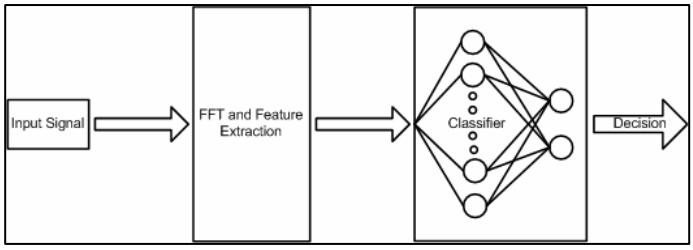

Fig. 1. Gender Classification system Architecture

2 Audio classifier

Our method is based on neural network for classification. Proposed method has 2 parts, after reading data from database tulips1 [7], first part is feature extraction and next part is our classifying based on neural network. Fig. 1 shows our system architecture. Next section describes all parts of our algorithm.

3 Feature extraction

Most important part in classification is feature extraction, because features determine differences between different signals and data. Main features are pitch and acoustic feature. These features are described in the following.

3.1 Pitch features

The pitch feature is perceptually and biologically proved as a good discrimin- ator between males' and females' voices. However the estimation of the pitch from the signal is not an easy task. Moreover, an overlap of the pitch values between male's and female's voices naturally exist, hence intrinsically limiting the capacity of the

pitch feature in the case of gender identification, Fig. 2 [1]. Hence, a major difference between male and female speech is the pitch.

In general, female speech has higher pitch (120 - 200 Hz) than male speech (60 - 120 Hz) and could therefore be used to discriminate between men and women if an accurate pitch [5]. By using auread command in MATLAB we read an .au file that consist voice of a male or female. With this command, we can convert an au file to a vector. For example, we read voice of a female in database (candace11e.au) and plot her audio signal in the following Fig. 3.

3.2. Acoustic features

Short term acoustic features describe the spectral components of the audio signal. Fast Fourier Transform can be used to extract the spectral components of the signal [1]. However, such features which are extracted at a short term basis (several ms) have a great variability for the male and female speech and captures phoneme like characteristics which is not required. For the problem of gender classification, we actually need features that do not capture the linguistic information such as words or phonemes.

4 The Classifier

The choice of a classifier for the gender classification problem in multimedia applications basically depends on the classification accuracy. Some of the important classifier is Gaussian Mixture Models (GMM), Multi Layer Perceptron (MLP), and Decision Tree. In similar training condition MLP has better accuracy in classification [1]. In this paper we used a MLP neural network for classifying, hence we describe a MLP in following, briefly.

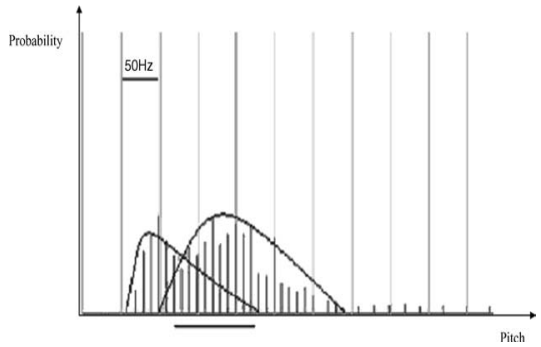

Fig. 2. Pitch Histogram for 1000 seconds of males (lower values) and 1000 seconds of females' speech (higher values). We can see the overlap between two classes.

4.1 Multi Layer Perceptron

MLP imposes no hypothesis on the distribution of the feature vectors. It tries to find a decision boundary, almost arbitrary, which is optimal for the discrimination

between the feature vectors. The main drawback for MLPs is that the training time can be very long. However, we assume that if the features are good discriminators between the classes and if their values are well normalized the training process will be fast enough.

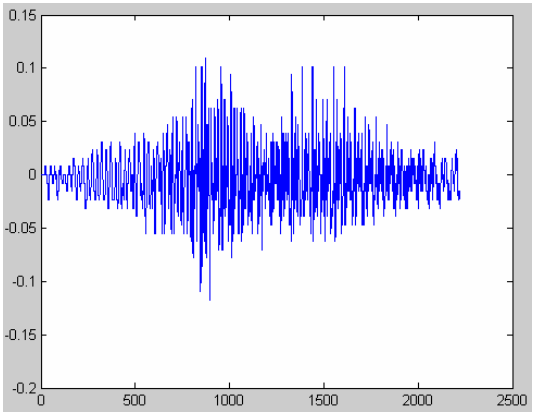

Fig. 3. A female audio signal that plot and show samples of this signal and their values between -0.2 and 0.2.

5 Proposed approach

In our method we processed audio signal that capture from a database (tulips1) contain 96 .au files. In this database every signal has a length about 1 second and we used some of this data for classifier training, and another files used for testing. First, we read 48 sound files that consists 3 males and 3 females, and with these data we train our network. As a classifier we use a multi layer perceptron with one hidden layer, 11 hidden neurons, and 2 output neurons that determine input vector is a male audio sample or female. For training an error backpropagation algorithm is used. First we used trainlm function for training, but for our application and with 1000 epochs for training, this function work very slow and it requires a lot of memory to run. Accordingly, we change Backpropagation network training function to TRAINRP. This function is a network training function that updates weight and bias values according to the resilient backpropagation algorithm (RPROP) and TRAINRP can train any network as long as its weight, net input, and transfer functions have derivative functions. Inputs data to this network are product of some preprocessing on raw data. Also, Transfer functions of layers in our network are default function in MATLAB (tansig). After reading data from database we get a Discrete Fourier Transform from input vectors by FFT(X, N) command. Fast Fourier Transform can be used to extract the spectral components of the signal. This command is the N-point FFT, padded with zeros if X has less than N points and truncated if it has more.

N in our problem is 4096, because with this number of point we can cover input data completely. After that, network training has been started with this vector as an input.

6 Experiments

The database used to evaluate our system consists of 96 samples with about 1 second length and we train our network with 50 percent of its data. Training data are consisting of 3 women' voices, 24 samples and three men, 24 samples (every person said one, two, three and four each of them twice). After training, we tested our classifier with another half of database and 96 % accuracy is obtained in gender classification.

7 Conclusion

The importance of accurate speech-based gender classification is rapidly increasing with the emergence of technologies which exploit gender information to enhance performance. This paper presented a voice-based gender classification system using a neural network as a classifier. With this classifier and by using pitch features we attained 96 % accuracy.

8 Future works

In the future, by using other features and using wavelet instead of Fourier transform or with that, we can get better results and achieve to higher performance. Also combining pitch and HMM for gender classification can be used to improve power of classification. And by dependent to problem, by using other classifier, better result may be obtained.

References

1. Hadi Harb, Liming Chen, Voice-Based Gender Identification in Multimedia Applications, Journal of Intelligent Information Systems, 24:2/3, 179–198, 2005.
2. Marston D., Gender Adapted Speech Coding, Proc 1998 IEEE International Conference on Acoustics, Speech, and Signal Processing, 1998. ICASSP 98, Vol. 1, 12–15, pp. 357 –360.
3. Konig, Y. and Morgan, N., GDNN a Gender Dependent Neural Network for Continuous Speech Recognition, International Joint Conference on Neural Networks, 1992. IJCNN, Vol. 2, 7–11, pp. 332–337.
4. Rivarol, V., Farhat, A., and O'Shaughnessy D., Robust Gender-Dependent Acoustic-Phonetic Modelling in Continuous Speech Recognition Based on a New

Automatic Male Female Classification, Proc. Fourth International Conference on Spoken Language, 1996. ICSLP 96, Vol. 2, 3–6, pp. 1081–1084.

5. Parris, E.S. and Carey, M. J., Language Independent Gender Identification, Proc IEEE ICASSP, pp. 685–688.
6. Martland, P., Whiteside, S.P., Beet, S.W., and Baghai-Ravary, Analysis of Ten Vowel Sounds Across Gender and Regional Cultural Accent Proc Fourth International Conference on Spoken Language, 1996. ICSLP 96, Vol. 4, 3–6, pp. 2231–2234.
7. Quast, Holger, Automatic Recognition of Nonverbal Speech: An Approach to Model the Perception of Para- and Extralinguistic Vocal Communication with Neural Networks, Machine Perception Lab Tech Report 2002/2. Institute for Neural Computation, UCSD. Download Website: http://mplab.ucsd.edu/databases/databases.html#orator