Classification of Speech Emotion Based on Gender Using Fuzzy Logic And Neural Network

Gowder Praveena Hiriyan
Assistant Professor,
KG College of Arts & Science, Coimbatore.

Abstract- An human interaction with computer become common interface now a days. Many developers identified Emotions from speech signal by classifying the emotion as happiness, surprise, neutral state, anger, sadness etc. The samples and the features are extracted from the emotional speech samples, sample taken as pitch. Another new technique utilizes fluffy rationale and neural system to distinguish the sex of the speaker. To prepare fluffy rationale and neural system, preparing informational index is created. The performances of the classification is based on the extracted features and the limit of the speech based on the classification.

Keywords- Neural network, Mel frequency Cepstrum Coefficient, Power spectrum.

I. INTRODUCTION

Emotion recognition improves the quality of human and computer interaction and easily interact with the system. The most natural way of interacting human with computer is speech by using speech tool.speech recognition is now good enough to allow speech to text engines, emotion recognition can increase the overall efficiency of interaction and may provide everyone a more comfortable user interface. It is often very difficult for humans to get the emotion of the speech signal and adjust their behavior accordingly. The arrangement of Emotion acknowledgment will allow to developer to build up a manmade brainpower that can get the speaker's sentiments that can be utilized as a part of numerous situations from PC diversions to virtual deals programs. Speech is one of the natural forms of communication. Emotional speech recognition is to detect and identifying the emotion from this emotions it identifies the gender by his or her voice by a use of new method fuzzy logic and neural network to identify the gender of the speaker. To train fuzzy logic and neural network, training data set is generated by using the above three features. At that point mean esteem is figured for the acquired outcome from fluffy rationale and neural system. By utilizing this limit esteem, the proposed technique recognizes the speaker has a place with which sexual orientation. The improved result shows the performance of the proposed technique in gender classification. Signal may be a normal or along with an

emotion like happy, sad, fear, angry etc. Voices differ for men and women in several aspects such as speaking pitch, pitch range, the space between the vocal folds, formant frequency, and the incidence of voice problems. Females speak with a higher fundamental frequency (voice pitch) when compared to males [1]. The higher pitch in women compared to men means the vocal folds vibrate or come together almost twice as many times per second in females than in males. The other differences found in voice quality are caused by the way the vocal folds vibrate between male and female. Usually males speak decrepit than female and females speak by causing an audible sound of breathing than males.

Many speech sets that belong to these emotion groups are taken and used for training and testing. The ERNN distinguishing these test samples. Neural networks are chosen for the solution because a basic formula cannot be devised for the problem. The neural networks are also quick to respond which is a requirement as the emotion should be identified almost instantly. The training takes a long time but is irrelevant as the training will be mostly off-line and on-line both.

II. GENDER CLASSIFICATION USING FUZZY LOGIC AND NEURAL NETWORK

This technique is used to identify the gender of the speaker. Various techniques are their for classifying the gender. Usually the major problem comes in pitch value detection. The pitch value depends upon the frequency of voice. Normally the pitch of male is low and for female the pitch will be high. Sometimes the pitch of male goes higher as the female and also the pitch of female goes lower as male. In such cases or situation speech classification using pitch will not produce approximate results. For this drawback here proposed a new method for speech classification using three features, energy entropy, short time energy, and zero crossing rates.

Basically the three feature values are calculated and given as an input to the fuzzy logic and neural network separately and it gives the percentage of male and the female feature as output. After that the mean esteem is taken for this

utilizing this esteem sexual orientation grouping. The process method is explained .Here discussed about the features which are used in our method. Speech Signals Feature analysed by selection plays one of the vital roles in gender classification. It depends on the feature which we have selected in proposed method. The three features used in our method are as follows:

•Short Time Energy (STE).
•Zero Crossing Rate (ZCR).
•Energy Entropy (EE).

2.1.1. STE

The STE of speech signal is said to be the sudden increase in energy signal. To compute STE, initially the signal is split into windows and then the window function is calculated for each window. The STE is calculated using the equation given below.

$$S = \sum_{r=-\infty}^{\infty} y(r)^2 . h(s-r) \qquad (1)$$

By using the above equation the STE is calculated. From the testing results we have observed that the energy entropy output for males is low whereas for females it is high and continuous.

2.1.2. ZCR

The ZCR is the most important feature considered in our method. The ZCR is defined as to be the ratio of number of time domain zero crossings occurred to the frame length. The equation 3 shows the formula to calculate zero crossing rate.

$$Z = \frac{1}{2N} \sum_{i=1}^{N-1} sgn\{x(i)\} - sgn\{x(i-1)\} \qquad (2)$$

$$sgn\{x(i)\} = \begin{cases} 1 ; x(i) > 0 \\ 0 ; x(i) = 0 \\ -1 ; x(i) < 0 \end{cases} \qquad (3)$$

By using the above equation the ZCR for each signal is calculated. From the testing comes about we watched that the ZCR for female discourse is higher than that of the male discourse. After age of fluffy standards the subsequent stage is to prepare fluffy rationale. The fuzzy logic is trained by using the rules shown in Table 1. To train fuzzy logic, training datasets are to be generated. The input training dataset is generated as {[Emax,Emin], [Smax,Smin], [Zmax,Zmin]}. After completion of training, the fuzzy logic obtained is ready for practical operation. Intesting if we will give E, S and Z

values as input to the fuzzy logic it will provide the output as the feature belong to male or female.

2.1.3. EE

EE in discourse flag is characterized as the sudden diverse changes in the vitality level of a discourse flag. To calculate EE, initially the speech signal is split into kframes and then the normalized energy for each frame is evaluated. The formula to calculate energy entropy is given below:

$$E = -\sum_{i=0}^{k-1} \sigma^2 . log_2(\sigma^2) \qquad (4)$$

where, σis the normalized energy. By using the above equation the EE is computed. From the testing results we have observed that the energy entropy for males is low and distributed while for females it is high and remains for a short period. The features used in our method are explained in the above sections. Next process is to identify the percentage of male and female feature present in the given speech signal using fuzzy logic and neural network.

2.2. Identifying Male and Female Feature using Fuzzy Logic Fuzzy Logic Offers several unique parameters which alternatively produces better results in many control problems [5]. Fuzzy logic here is used to calculate the percentage of various male and female features presents in the given speech signal. Generally fuzzy logic consists of three important steps. This includes fuzzification, generating fuzzy rules and defuzzification. In the fuzzification process the system data is converted in to fuzzy data. For fuzzification process triangular membership function is used. Next process after this is generating fuzzy rules. Figure1 shows the structure of fuzzy logic used in the proposed method with 3 input variables and one output variable.
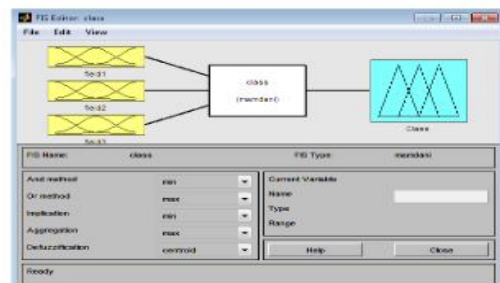

Figure 1. Structure of fuzzy used in our method.

2.2.1. Fuzzy Rules Generation

The contribution to our fluffy rationale is vitality entropy (E),short time vitality (S) and zero intersection rate (Z) and the yield got from the fluffy is the level of different male and female highlights which are available in the given discourse flag. The input variables are fuzzified into three various sets namely; large, medium and small and the output variable is fuzzified into three sets namely; male, female/male and female. In female/male the discourse flag has a place with either male or female. The fluffy tenets produced are appeared in Table1.

Table 1. Fuzzy rules.

S. No	Fuzzy Rules for Gender Classification
1	if E=high and S=low and Z=low, then Male
2	if E=high and S=low and Z=medium, then Female/Male
3	if E=high and S=low and Z=high, then Female
4	if E=high and S=medium and Z=low, then Female/Male
5	if E=high and S=medium and Z=medium, then Female
6	if E=high and S=medium and Z=high, then Female
7	if E=high and S=high and Z=low, then Female
8	if E=high and S=high and Z=medium, then Female
9	if E=high and S=high and Z=high, then Female
10	if E=medium and S=low and Z=low, then Male
11	if E=medium and S=low and Z=medium, then Male
12	if E=medium and S=low and Z=high, then Female
13	if E=medium and S=medium and Z=low, then Female/Male
14	if E=medium and S=medium and Z=medium, then Female/Male
15	if E=medium and S=medium and Z=high, then Female/Male

3.1. Identifying Male and Female Feature using Neural Network

The main aim of the classification ANNs is to produce an exact output based on the input parameters [7]. Neural networks are used here to calculate the percentage of female and male features present in agiven speech signal. Basically neural network consists of three layers namely; input layer, hidden layer and output layer. In our method input layer has three variables, hidden layer has n variables and output layer has one variable. The input to the neural network is energy entropy, short time energy and zero crossingrate. The two stages of operation which takes placein neural network are training stage andtesting stage. For training of neural network, training dataset is generated. The input training dataset is generated as

{[Emax,Emin],[Smax,Smin],[Zmax,Zmin. Figure 2 shows the structure of neural network used in the proposed method with 3 input variables and one output variable.
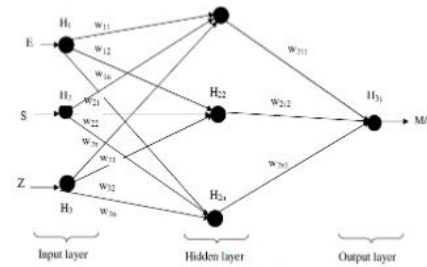

Figure 2. Structure of neural network used in proposed method.

3.1.1. Neural Network Training for Gender Classification

The steps for training the neural network are:

•Step 1:Initialize the input weight of each neuron.
•Step 2:Apply a training dataset to the network.
Here E, Sand Zare the input to the network and M/Fis the output of the network.

$$M / F = \sum_{r=1}^{n} W_{2r1} y(r) \qquad (5)$$

where,

$$y(r) = \frac{1}{1 + exp(-w_{11r}.(E + S + Z))} \qquad (6)$$

Equation 5, represents the activation function performed in the output and input layer respectively.

•Step 3:Adjust the weights of all neurons.
•Step 4:For each E, Sand Z corresponding male and female feature is computed.
•Step 5:Repeat the iteration process till the output reaches its least value. After completing the training neural network is ready for various practical applications. Next step after completion of training is testing neural network. Amid testing discourse flag is given as information, it gives the level of male and female element show in that flag.

3.2. Gender Classification for the Given Speech Signal After completing the process of training fuzzy logic and neural network, the next process is to identify the gender of the speaker. The initial step is to compute the mean value of output obtained from fuzzy logic and neural network. The mean value is calculated using the equations given below:

$$S^{final} = \frac{S^{fuzzy} + S^{NN}}{2} \qquad (7)$$

where,

S fuzzy is the output generated from fuzzy logic and SNN is the output obtained from neural network. After calculating these mean values, the speech signal is splitted in to male and female using a threshold value.

$$classification = \begin{cases} if\ S^{final} < S^{threshold}\ ; then\ female \\ f\ S^{final} > S^{threshold}\ ; then\ male \end{cases} \quad (8)$$

From the above equation we obtain the speaker belongs to which gender. The threshold used in our method is 0.5.

IV. RESULT AND DISCUSSIONS

This proposed technique was implemented in MATLAB 7.10 and is tested for different speech signals from Harvard-Haskins database [24]. Here 80speech signals are taken as an input and then splitted into four datasets. Initially the neural network and fuzzy logic is trained by using some speech signal,and testing is performed by using a set of speech signal as input to the proposed method so that it identifies the speaker gender. The result of proposed technique i.e., combination of fuzzy logic and neural network are compared with the Fuzzy Logic (FL) and Neural Network (NN), Naive Bayes (NB) and using pitch as feature. From the comparison results, it is clear that our method is better than the other methods. The performance of the proposed method, fuzzy logic and neural network are explained separately in the below sections.

4.1. Performance Analysis of Gender

Classification The True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) values are calculated from the results obtained to the proposed method, fuzzy logic and neural network. The above four values are used to compute performance parameters like false positive rate (α), false negative rate (β), sensitivity (SE), specificity (SP), Likelihood Ratio Positive (LRP), Likelihood Ratio Negative (LRN), Accuracy (Acc) and Precision (Pre) using the equations given below:

$$SP = \frac{TN}{(FP+TN)} \quad (9)$$

$$SE = \frac{TP}{(TP+FN)} \quad (10)$$

$$\alpha = \frac{FP}{(FP+TN)} \quad (11)$$

$$\beta = \frac{FN}{(TP+FN)} \quad (12)$$

$$LRP = \frac{Sensitivity}{(1-specificity)} \quad (13)$$

$$LRN = \frac{(1-Sensitivity)}{specificity} \quad (14)$$

Using these above equations the performance of proposed method, FL, NN, NB and using pitch is calculated and values obtained.

V. CONCLUSIONS

In this paper, a gender classification technique in speech processing using neural network and fuzzy logic was proposed. In this technique gender classification is performed by considering three different features such as energy entropy, short time energy and zero crossing rates. Firstly mean values are calculated for three features by using training dataset and percentage of male and female features which are present in the speech signal are computed using fuzzy logic and neural network individually and then the mean value is taken to identify the gender of the speaker. This approach was implemented in the working platform of MATLAB for testing. The proposed method was tested using Harvard-Haskins database. During testing if a speech signal is given as input it will identify the gender of the speaker to which speaker belongs. The results obtained from proposed method are compared with the fuzzy and neural network, using pitch as feature. Comparison results have shown that our method is better than the other methods in gender classification.

REFERENCE

[1] Devi M., Kasthuri N., and Natarajan A., "Performance Comparison of Noise Classification usingIntelligent Networks," International Journal of Electronics Engineering, vol. 2, no. 1, pp. 49-54, 2010.

[2] Gomathy M., Meena K., and Subramaniam K., "Gender Grouping in Speech Recognition using Statistical Metrics of PitchStrength,"
European Journal of ScientificResearch, vol. 61, no. 4, pp. 524, 2011.

[3] Gudi A., Shreedhar H., and Nagaraj H., "Signal Processing Techniques to Estimate the Speech Disability

in Children," IACSIT International Journal of Engineering and Technology, vol. 2, no. 2, pp. 169-176, 2010.

[4] Gudi A. and Nagaraj H.,"Optimal Curve Fitting of Speech Signal for Disabled Children," International Journal of Computer Science &Information Technology, vol. 1, no. 2, pp. 99-107, 2009.

[5] Haider T. and Yusuf M., "A Fuzzy Approach to Energy Optimized Routing for Wireless Sensor Network," The International Arab Journal of Information Technology, vol. 6, no. 2, pp. 179-188, 2009.

[6] Emotion Recognition Using Neural Networks MEHMET S. UNLUTURK, KAYA OGUZ, COSKUN ATAY Proceedings of the 10th WSEAS International Conference on NEURAL NETWORKS 2010.

[7] DWT and MFCC Based Human Emotional Speech Classification Using LDA M Murugappan, Nurul Qasturi Idayu Baharuddin, Jerritta S 2012 International Conference on Biomedical Engineering (ICoBE),27-28 February 2012,Penang.

[8] J. A. Freeman and D. M. Skapura, Neural Networks: Algorithms, Applications and Programming Techniques, Addison Wesley Publishing Company, 1991.

[9] T. Masters, Practical Neural Network Recipes in C++, Academic Press Publishing Company, 1993.

[10] A. Cichocki and R. Unbehauen, Neural Networks for Optimazing and Signal Processing, John Wiley & Sons Publishing Company, 1993.