2012 International Conference on Advanced Computer Science Applications and Technologies

Speech Recognition by Classifying Speech Signals Based on the fire fly and fuzzy

Fatemeh Hoseinkhani[1]
Department of Electrical Computer and Biomedical Engineering,
Qazvin Branch Islamic Azad University
Qazvin,Iran
fatemeh_hoseinkhani@yahoo.com

Ebrahim Parcham [2]
Department of Electrical Computer and Biomedical Engineering,
Qazvin Branch Islamic Azad University
Qazvin,Iran
parcham@live.com

monireh pournazari[3]
Department of Electrical Computer and Biomedical Engineering,
Qazvin Branch Islamic Azad University
Qazvin,Iran
m_pournazary@yahoo.com

Nasim Borzue[4]
Department of Electrical Computer and Biomedical Engineering,
Qazvin Branch Islamic Azad University
Qazvin,Iran
N_Borzue @yahoo.com

Abstract—**In this paper, a fuzzy speech recognition system is proposed for classifying speech signals to get more accurate recognition with higher speed. This system is combined with 5-layers fuzzy logic and is more accurate in speech recognition than other methods like: particle swarm optimization–Forward Neural Network (PSO-FNN) and Back Propagation Forward Neural Network (BP-FNN). In this paper, speech samples are first given to input of the fuzzy circuit to check (investigate) signals in a fuzzy framework and a pattern of signals is produced for each signal cluster. This causes dimension reduction of signal data and gives us better and more reliable recognition result. For recognizing speech, we use firefly classification method and consider a special class for each input to improve recognition rate. Classifying fuzzy signal is the reason for increasing recognition accurancy. Our method is also capable of recognizing noises in environment around and consider each noise individually as a cluster and then removing it from input signal for final recognition. Our classification method based on firefly algorithm improves recognition of speech signals in the proposed model.**

Keywords : Voice recognition – Fuzzy logic – Classification – Firefly – Denoising.

I. Introduction

Speech recognition has become the most important issue in signal processing during two previous decades and lots of methods are proposed (taken in to consideration) for this issue. Each of them containing their own advantages and dis advantages. For example methods like PSO-FNN and PB-FNN was not able to recognize similar signals and caused fewer comparisons, for similar signals by extracting signal features quickly. Besides, noises of around environment used to destroy the features of that signal. One of disadvantages of fuzzy neural networks is that they encounter enormous amount of speech signals and this takes processing time. Thus, optimal order of neural network is significant in speech recognition to have suitable networks weights. In this paper we propose a new method based on fuzzy neural network (FNN) classification that resolves disadvantages of previous methods such as: low processing speed and problems of recognizing properties of similar signals. This method works with higher speed and accuracy in speech recognition. Moreover, this method has a suitable pattern for denoising environment noises and do this by considering testing patterns and sensitive edges of words signals. The high speed of speech recognition is because of minimizing range of signals sent via fuzzy circuit. The high accurancy in speech recognition and also discriminative signals is because of firefly clustering. However, by creating noisy classes for each speech, we will be able to recognize speech in noisy conditions and denoising capability is performed easily on original signals. In previous methods for speech recognition, the orginal signal itself was used for noise reduction and signal might be damaged due to false recognition of noise. In our proposed method, after probabilistic speech recognition, the input signal is compared with trained signals and the noise of this signal is reduced by noisy classes and noiseless trained signals if the input signal and the trained signals are similar.

We also perform speech recognition parallel with checking the output of fuzzy circuit and compare closed classes. In this way the accurancy of our method improves speech recognition. In speech recognition, we also face uncertainty of words that is illustrated in different cases. To fix this problem we use a 5-layered fuzzy circuit. Classification helps us to recognize speech by considering closer class to it. This increases the speed of speech recognition process and makes it more accurate by comparing the output of fuzzy circuit.**[1-2-9-11-12]**

978-0-7695-4959-0/13 $25.00 © 2013 IEEE
DOI 10.1109/ACSAT.2012.31

II. DESIGNING FUZZY PART

Our proposed method for speech recognition contains 5-layers as shown in figure1. x_i denotes the input speech signals and y_i denotes the output of the circuit. As it can be seen, the outputs are of smaller sizes than the inputs. The length of this outputs depends on rules created in fuzzy circuit. These rules have different sizes on the basis of different signals. The outputs devote a value to each input speech signal in continuous mode as a key or signal pointer [1,2,3].

Fuzzification of inputs:

Equation (1) implies triangular membership function in which a ,b , and c are locations of putting triangular function on signal x. In this equation, fuzzy inference engine receives inputs and determines the degree of inputs for each of fuzzy sets: [8-2]

$$u(x;a,b,c) = \max\left(\min\left(\frac{x-a}{b-a},\frac{c-x}{c-b}\right),o\right) \qquad (1)$$

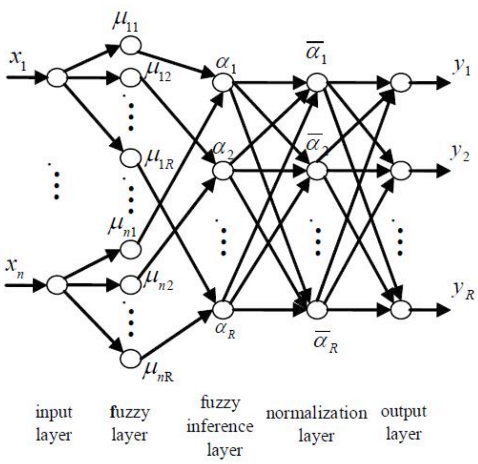

Figure 1.Fuzzy Diagram.

Applying fuzzy operators:

After fuzzification, the true degree of each input component is determined. Equation (2) is the function computing fuzzy interface:

$$j = \prod_{i=1}^{R} \mu_{ij} * \frac{1}{N_{adj}}\alpha \qquad (2)$$

in which R refers to the number of some parts of fuzzy operators that is created for combination and the number implying degree of parts is provided by it. N_{adj} is the equalization coefficient that we consider it as n/4 is this paper.

Applying implication method:

The result is a fuzzy set determined by membership functions. Input of this process is a number (digit) and the output is a fuzzy set. The output of node J is equivalent to equation(3):

$$\bar{\alpha}_j = \alpha_j \,/\, \sum_{i=1}^{R} \alpha_i \qquad (3)$$

Summation of output:

Since in a fuzzy interface system, decisions are made by evaluating all rules, in this layer we combine these rules and the output of node K is as below:

$$y_k = \sum_{j=1}^{R} w_{jk}\bar{\alpha}_j \;, k = 1,2,A,R \qquad (4)$$

Here w_{jk} is the weight of each rule that is applied on the obtained value of presumption part. After creating output signals, we have to classify them to gain better recognition. To do so we take advantage of firefly algorithm. Firefly classification method is more speedy than the other clustering methods. We will show rate of accurancy in the following of this paper.[12-2-3]

III. CLUSTERING FUZZY OUTPUT SIGNALS BY MEANS OF FIREFLY ALGORITHM

A. Clustering

Clustering is an important unsupervised classification technique in which a set of patterns (usually vectors in a multi-dimensional space) are classified in to some criteria such as: euclidean distance, mahalanobis distance, chernoff distance and so on. [6-7]

Clustering algorithms are divided into two parts: Heirarchical clustering and partitioning clustering, Heirarchical clustering is a kind of clustering with heirarchical structure of clusters built by dividing a big cluster into small clusters and then merging these small clusters by considering the closet obtained center of gravity. Here there are two types of Heirarchical clustering methods as bellow:

- Divisive method, that divides a big cluster into two or more smaller clusters.
- Agglomerative method in which a big cluster is built by merging two or more smaller clusters.

On the other hand, Partitioning clustering tries to divide the Dataset into a discrete cluster set without hierarchical structure. In most cases, the used partioninig clustering algorithms have applied clustering algorithms based on original samples where each cluster is displayed by its center. Target function (Error Square Function) is the summation of patterns distances from center. In this paper we are mostly interested in Partitioning clustering to produce cluster centers and then to classify DataSets by means of this cluster. [4-5-7]

B. Firefly algorithm

Fireflies are kinds of insects that illuminate their environment around. For better illustration of this algorithm, we use three desired rules: [10]

- All Fireflies are similar to each other. one Firefly attracts other Fireflies without paying attention to their sexuality.
- An important and interesting behavior of fireflies is that the one lights better, will own prey and will share it with others.
- Lighting is an important factor for Fireflies attractions. So Firefly agent moves toward its own neighbor which is more lighting.

Firefly algorithm (FA) is a population-based algorithm. This algorithm exports for global optimum of target functions based on swarm intelligence to evaluate heuristic behavior of fireflies. In FA, physical entities (Agents or fireflies) distribute randomly over problem space. Agents are known as fireflies and the quality of lighting is called light intensity. Each of fireflies is attracted by its own brighter neighbors and their attraction is decreased by increasing the distance between them. If non of fireflies are brighter than the other, then they would move randomly. In application of FA clustering determinant variables are cluster centers. Target function is sum of euclidean distance between all training data samples in N dimensional space. All agents (fireflies) are spreaded randomly through search space based on this function and also being initialized. Two phases of firefly algorithm is as follows:

- Illumination variation: it is related to target valus so that in maximization (minimization) problems, a firefly attracts high (low) illumination intensity according to high (low) illumination intensity of the other firefly. Suppose we have a n-Agent group of fireflies and x_i is a solution for firefly i that $f(x_i)$ shows its fitness value. Here I_i is the illumination intensity of a selected firefly to reflex recent location of X and fitness value of f(x).

$$I_i = f(x_i) \qquad 1 \le i \le n. \tag{5}$$

- Moving toward more attractive firefly: an attractive firefly radiates an illumination intensity which is obvious to other fireflies. Each firefly has a definite attraction β which determines how much that firefly is strong in attracting other group members. However, fireflies attraction (β) is related to r_{ij} that denotes the distance between two fireflies i and j which are located in places x_i, x_j . The distance r_{ij} is defined as follows: [19-20]

$$r_{ij} = \|x_i - x_j\| \tag{6}$$

Attraction function of a firefly $\beta(r)$ is determined by relation below:

$$\beta(r) = \beta_0 e^{-\gamma r^2} \tag{7}$$

In which β_0 is the attraction value in r=0, and γ is called illumination absorbtion coefficient.

Movement of firefly i that is located in place x_i, toward brighter firefly j that is located in place x_j is shown by equation(8):

$$x_i(t+1) = x_i(t) + \beta_0 e^{-\gamma r^2}(x_j - x_i) \tag{8}$$

C. Firefly clustering algorithm

Clustering methods are developed on unsupervised learning method without noticing the objects in groups or classes. In unsupervised technique, training sets of data are classified in data (such as clusters centers) based on numeric information and then are matched via information class analyzer. Datasets which we follow contain class information for each data. So, the main target of finding clusters centers is to minimize target function (sum of pattern distances from clusters centers). For N objects in a given problem, the target is to minimized sum of euclidean distance between all patterns and devoting each of patterns to one of K clusters. Clustering target function for sum of square error is computed by relation (9):

$$\mathrm{J(K)} = \sum_{k=1}^{k} \sum_{i \in c_k} (x_i - c_k) \tag{9}$$

In relation (9) K is the number of clusters for n patterns (i=1,2,……,n), x_i is the place of pattern i and c_k (c=1,2,…….,k) is kth cluster center that is computed by relation (10):

$$c_k = \sum_{i \in c_k} \frac{x_i}{n_k} \tag{10}$$

In equation (10), n_k is the number of patterns in cluster K.

Analyzing clusters is in the way that dataset is devoted to clusters which patterns are classified in to one cluster based on some similarity criteria.

To evaluate the distance between patterns similarity measurement criteria are usually used. Clusters centers are decision making variables which are obtained by minimizing sum of euclidcan distance between all training samples in n_dimensional space.

Target function for pattern i is computed as bellow:

$$f_i = \frac{1}{D_{Train}} \sum_{j=1}^{D_{Train}} d(x_j, p_i^{CL_{known(x_j)}}) \qquad (11)$$

In equation(11), D_{Train} is the number of training datasets which are used to normalize the sum and is located in range [0.0,1.0], and $p_i^{CL_{known(x_j)}}$ is defined as a class to which samples similar to database belong.

Note that in FA algorithm, decision making variables are clusters centers. Target function in firefly algorithm is determined by equation(11). For a dataset: n is the number of data points, d denotes problems dimension and c is the number of classes. A data point belongs only to one of c classes.

D. Noise recognition in clustering

In this clustering method, we have noise in classes that causes decrease of accurancy, since in this clustering method each signal is located in a special class and noise is an unremoved agent in environment. To recognize a special pattern of noise and to decrease errors in speech recognition, we consider mean of pronouncation differences of a word as the input pattern of circuit and store it in a special class that is called noisy class. The more the noisy class is, the better noise recognition in environment is achieved. For example, suppose pronunciation of digit 3, note that signals are obtained in 12 different cases with different noises. As shown in figure 2, the pronounced word is obvious in begin and end of signal and also in the word's signal. [13-14-17]

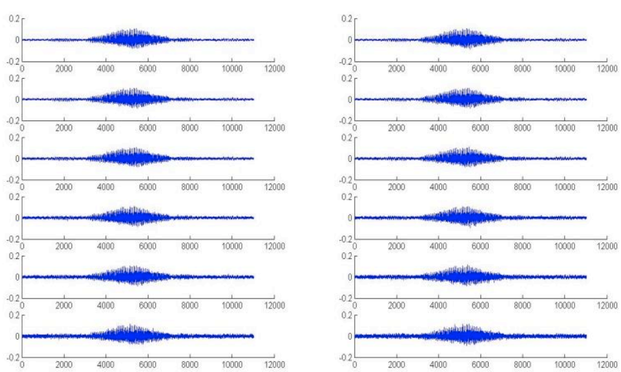

Figure 2. Sample Pronunciation Of Digit 3 in different white noises.

In signal denoising part, we first find mean of difference between noiseless signals and noisy signals. This difference determines the amout of signal noise and then it is given to circuit as a noisy class sample to be stored in one of the word's noisy classes. In program test, the pronounced word (digit 3) is compared with dataset gathered in section 4 and then the noisy class of this signal is fetched and subtracted by input signal in frequency domain. The achieved signal is compared with orginal signal and if they are equivalent or have less differences then the noise of signal is removed, otherwise it is considered as a new noisy signal and the mean of that signal is added to previous mean of signals and another noisy class is created. In figure a-3, you can observe background noise is completely removed and the quality and details of speech is not changed. [15-19]

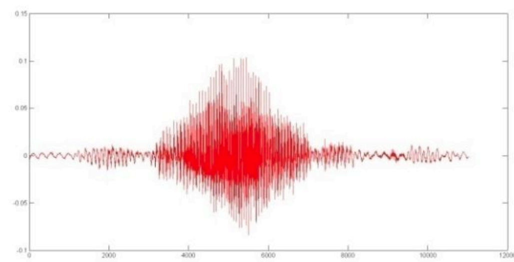

Figure 3-a. Initial Signal.

Fireflies are initialized in search space randomly. Values of used parameters in algorithm is as follows:

Number of fireflies N is equivalent to 20. Value of initial attraction is equivalent to 1. illumination attraction coefficient $\gamma$is equivalent to 1. Number of generations T is equivalent to 100. [16]

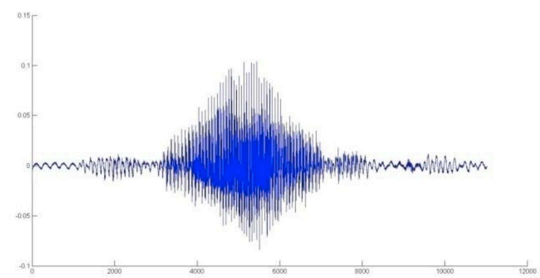

Figure 3-b . The Correct Signal With Noisy Classification

E. FA clustering and parameter setting

IV. EXPERIMENTAL RESULTS

In this paper, we have used 45 people containing 25 woman and 20 men to create our DataSet. For every person, we performed this work by considering the pronunciation of 30 Farsi words (digit between 1 and 50) and each word is pronounced in 6 different way (different accents). Sampling is done in 11.025 speed frequency. This algorithm is written in presence of white Gaussian with the rate of signal to input noise: 15 db, 20 db, 25 db and 30 db and also a noiseless model. To execute the proposed algorithm, after giving speech signals in to fuzzy circuit and achieving output signals, we perform clustering. This clustering for speech is put in a set of pronounced digits. Then we consider a noisy signal class for each speech of each person. Inspecting a talker's speech, a special class is created for that talker and is compared by other classess to select the nearest class. Then for more accurate signal recognition, we compare the fuzzified outputs of fuzzy circuit with the output of found classes, and if two signals are similar, we check noise within this signal and remove the noise by method explained in section 4-3 then we update noise of noisy classes. In table 1, one of outputs of proposed algorithm is compared with other algorithms based on given noisy models and the effect of FAFC-FNN is implied in increasing accurancy and efficiency in speech recognition.

V. CONCLUSION

In this paper, we have used a hybrid method for speech recognition that is a combination of forward Neural Networks and clustering fuzzy output signals based on firefly algorithm. In this method, we have given input signals to fuzzy system and cluster these signals for better speech recognition. Clustering processs is done by means of firefly algorithm. Then we denoise these signals with achieved noisy classes while clustering to recognize final speech signal noise class. This hybrid method has some advantages such as: high speech in processing because of dimension reduction in checking signal in 5-layers fuzzy circuit, high speech recognition accurancy because of using clustering of output signals of fuzzy circuit and improvement in comparing output class with other speech classes, better and more accurate recognition and comparison between output class and outputs of fuzzy circuit from start class, resistance of this system against natural noises in environment around is because of putting noise in a seperated class for easier recognition. [18]

Table1. voice recognition and comparison between FAFC/FNN- PSO/FNN-BP/FNN Algorithms in different white noises

Vocabulary	SNR(dB) Training Algorithm	Comparison of different methods **FAFC**-FNN , PSO-FNN, BP-FNN in different white noises				clean	Maximum Difference	Average Recognition Rate
		15	20	25	30			
10	**FAFC**-FNN	97	96	97	97	97.9	26	96.98
	PSO-FNN	94.8	95.7	96.7	96.2	96.7	1.9	96.0
	BP-FNN	95.2	94.8	94.8	94.8	95.2	0.4	95.0
20	**FAFC**-FNN	96.2	96.7	97.1	96.4	97.6	1.4	96.8
	PSO-FNN	96.4	95.5	96.4	95.7	97.6	2.1	96.3
	BP-FNN	94.5	95.2	95.5	96.4	96.7	2.2	95.7
30	**FAFC**-FNN	96.1	96.4	97.3	97.6	97.9	1.8	97.1
	PSO-FNN	95.6	96.3	97.3	97.5	97.6	1.9	96.9
	BP-FNN	93.2	94.3	96.1	94.8	96.8	3.6	95.0
40	**FAFC**-FNN	96.1	97.1	97.7	97.9	98.3	2.2	97.3
	PSO-FNN	95.5	96.8	97.6	97.5	97.9	2.4	97.0
	BP-FNN	92.6	93.3	95.5	95.9	95.6	3.0	94.6
50	**FAFC**-FNN	95.8	97.1	96.8	97.3	97.8	2.0	97.0
	PSO-FNN	95	96.7	96.6	97.3	97.8	2.8	96.7
	BP-FNN	92.2	93.3	94.9	94.2	95.2	3.0	94.0

REFERENCES

[1] Gaoyun Li, “The Study of the Speech Recognition Based on theOptimized RBF Neural Network”, Master Degree Thesis of TaiyuanUniversity of Technology, Taiyuan, 2009.

[2] XueYingZhang ,Peng Wang, “Improved T-S Fuzzy NeuralNetwork in Application of Speech Recognition System”, ComputerEngineering and Applications, vol. 45, April 2009, pp 246-248.

[3] Yang Wu, Wu Yuan, Wu Zhongru, “Forecast Model Study Based onFuzzy Neural Network” ,Water Resources and Power, vol. 22, pp 63-65, Mar, 2004.

[4] M. SenthilArumugam, M.V.C.Rao, Alan W.C.Tan, “A novel andeffective particle swarm optimization like algorithm withextrapolation 2009.

[5].Whitehead ML, Stagner BB, Lonsbury-MartinBL, Martin GK. Measurement of otoacoustic emissions for hearing assessment. IEEE EngMedBiol Mag 1994;13(2):210-16.

[6] Otoacoustic emissions (OAEs) Portal [Online]2006 [2006 Nov]; AvailblefromURL:http://www.oae.it.

[7]Abolhasani MD, Salimpour Y, Sarkar S. Editors.Reproducibility enhancement of otoacousticemission based on multirate signal processing.Proceeding of Xth Mediterranean Conferenceon Medical and Biological Engineering;2004 jul 31-Aug 5; Ischina, Italy.

[8] Xiucheng Dong, Yunyuan Zhao, YunyunXu, Zhang Zhang and Peng Shi. Design of pso fuzzy neural network control for ball and plate system. International Journal of InnovativeComputing, Information and Control, 2011, pp. 7091-7103 .

[9] Jing-RuZhang , Jun Zhang , Tat-Ming Lok , Michael R. Lyu . A hybrid particle swarm optimization–back-propagation algorithm for feedforward neural network training. Elsevier , Applied Mathematics and Computation 185 (2007) 1026–1037.

[10] Y. Marinakis, M. Marinaki, M. Doumpos, N. Matsatsinis, C. Zopounidis, Ahybrid stochastic genetic-GRASP algorithm for clustering analysis, OperationalResearch An International Journal 8 (1) (2008) 33–46.

[11] A.K. Jain, M.N. Murty, P.J. Flynn, Data clustering: a review, ACM ComputingSurveys 31 (3) (1999) 264–323.

[12] Wei Shen , Xiaopen Guo , Chao Wu , Desheng Wu, Forecasting stock indices using radial basis function neural networks optimized by artificial fish swarm algorithm, Knowledge-Based Systems, v.24 n.3, p.378-385, April, 2011 [doi>10.1016/j.knosys.2010.11.001]

[13] XueYingZhang ,Peng Wang, “Improved T-S Fuzzy NeuralNetwork in Application of Speech Recognition System”, ComputerEngineering and Applications, vol. 45, pp 246-248, April, 2009.

[14] Yang Wu, Wu Yuan, Wu Zhongru, “Forecast Model Study Based onFuzzy Neural Network” ,Water Resources and Power, vol. 22, pp 63-65, Mar, 2004.

[15] Legchenko, A., Valla,P., " A review of the basic principles for proton magnetic resonancesounding measurements " J. APPL. Geophysics. 2002 , 50 , 3-19.

[16] Plata, J.L., Rubio, F.M., " Basic theory of the Magnetic Resonance Sounding method"BoletinGeologicoyminero, July –September 2007 , vol.118, 441-457.

[17] Legchenko, A., " MRS measurements and inversion in presence of EM noise" BoletinGeologicoyminero, July –September 2007 , vol.118,489-508.

[18] Plata, J.L., Rubio,F.M., " MRS experiments in a noisy area of a detrital aquifer in the southof spain" J.APPL. Geophysics,2002 , 50 , 83-94.

[19] Mallat,S.G., “A theory for multiresolution signal decomposition: The waveletrepresentation,” IEEE Transactions on pattern analysis & machine intelligence,July 1989, vol.11, pp. 674–693.

[20] Zhang J. R., Zhang J., Loc T. M., Lyu M. R., “A hybrid particle swarm optimization-backpropagationalgorithm for feedforward neural network training”, Applied Mathematics andComputation, 185, pp.1026-1037, 2007.