Speaker Verification System using a Hierarchical Adaptive Network-based Fuzzy Inference Systems (HANFIS)

Zohreh Soozanchi-K.[1] , Mohammad-R. Akbarzadeh-T.[2], Mahdi Yaghoobi[3], Saeed Rahati[4]

1- Department of Artificial Intelligence, Islamic Azad University, Mashhad Branch, Iran
2- Center for Applied Research on Soft Computing and Intelligent Systems, Ferdowsi University of Mashhad, Iran
3, 4- Department of Artificial Intelligence, Islamic Azad University, Mashhad Branch, Iran
Email: z.soozanchi@gmail.com

Abstract. We propose the use of a hierarchical adaptive network-based fuzzy inference system (HANFIS) for automated speaker verification of Persian speakers from their English pronunciation of words. The proposed method uses three classes of sound properties consisting of linear prediction coefficients (LPC), word time- length, intensity and pitch, as well as frequency properties from FFT analysis. Actual audio data is collected from fourteen Persian speakers who spoke English. False acceptance ratio and false rejection ratio as are evaluated for various HANFIS trained with different radius. Results indicate that vowel sounds can be a good indicator for more accurate speaker verification. Finally, the hierarchical architecture is shown to considerably improve performance than ANFIS.

Keywords: ANFIS, Speaker verification, LPC, FFT, intensity and pitch coefficients, HANFIS

1 Introduction

The computer industry controls most informational, financial and security systems. Currently, the research community has considered properties of face, sound, fingerprint, and iris for new identification schemes [1]. In addition to above approaches, sound is an important information source, is very simple for the user, and can potentially provide high speed recognition. To date, sound recognition for security purposes and with the use of ANFIS in particular, has been largely neglected. Few articles have applied neural networks and considered sound samples directly as input to the network, without any pre-processing. In this paper, we advocate the use of a hierarchical ANFIS (HANFIS) for fast clustering and identification of sound. Initially, several features are produced from sound of individuals after omission of noise and silence [2], and validation function is accomplished with these features by ANFIS. The experimental result of this paper which is practical represented by final Table at the end and it is MATLAB programme.

2 Sound Verification System

It is well known that people can be generally identified by their voices, not by the message sent by them. Basically, there are two kinds of voice based recognition or speaker recognition: speaker identification and speaker verification [3]. Speaker recognition is further divided into two categories, which are text dependent and text-independent [4]. The method that use in this paper is based on the concept of speaker verification since the objective in the access control is to accept or reject a person to enter a specific building or room. In general the speaker verification system for accessing control makes four possible decisions [5]. The accuracy of the access control system is then specified based on the rate in which the system makes decision to reject the authorized person and to accept the unauthorized person. The quantities to measure the rate of the access control accuracy to reject the authorized person is then called as false rejection rate (FRR) and that to measure the rate of access control to accept the unauthorized person is called to as false acceptance rate (FAR). Mathematically, both rates are expressed as percentage using the following simple calculations [3]:

$$FRR=NFR/NAA*100\% \quad (1)$$

$$FAR=NFA/NIA*100\% \quad (2)$$

NFR and NFA are the numbers of false rejections and false acceptance respectively, while NAA and NIA are the number of the authorized person attempts and the numbers of impostor person attempts. For achieving high security of the door access control system, it is expected that the proposed system will have both low FRR and low FAR. In order to give a definite answer of access acceptance or rejection, a threshold is set. When degree of similarity between a given voice and the model is greater than threshold, the system will accept the access; otherwise the system will reject the person to access the building/room.

3 Plan Executions by Hierarchical ANFIS

ANFIS, proposed by Jang [6], is an architecture which functionally integrates the interpretability of a fuzzy inference system with adaptability of a neural network. ANFIS structure is a weightless multi-layer array of five different elements [7].

In this research ANFIS-based speaker model is developed using Fuzzy Logic Toolbox of MATLAB and design of the ANFIS structure is done by determining premise parameters. Here the subtractive clustering method is used with different radius parameters. Once the premise parameters are obtained, the ANFIS model is trained by using hybrid learning algorithm.

In this case, we can use some ANFIS orders as shown in Fig. 1. We propose the use of a hierarchical adaptive network-based fuzzy inference system (HANFIS) for automated speaker verification of Persian speakers from their English pronunciation of words. In this way, we should allocate whole responses from each ANFIS as new recent data to other ANFIS input for its instruction. Here, we give answer to test

process of HANFIS with different pitch and intensity features, frequency features and LPC coefficients as input to the other network, then evaluate program with new data. As it was mentioned, we can obtain better results with the change of parameters and ANFIS's radius. Here, the percent of errors is zero with the change of radius. You can see combination of LPC features, frequency, pitch and intensity features in Table 4. The main important point in here is that the value of FAR should be close to zero, until unauthorized recognized better.

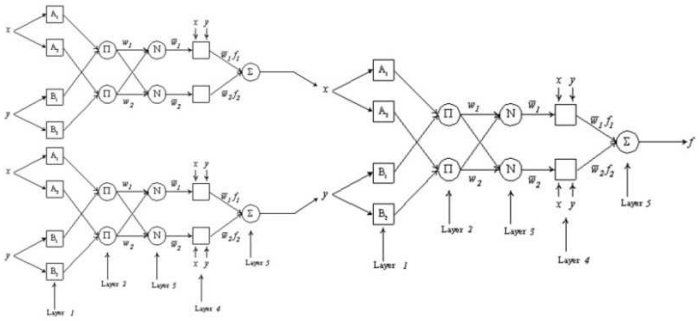

Fig.1. HANFIS architecture

4 Data Collection

Whole used data in this article is statement of word in English language by Persian speaker. At first, for gathering data by speaker, 20 samples from considered words were stated and recorded, then the silence between pronounced words at each sentence was omitted by PRAAT software and stored at different separated files (PRAAT is a program for doing phonetic analyses and sound manipulations. It can be downloaded from www.praat.org.). During arrangement, it is necessary to equalize length of whole samples because signal length in different speeches of fixed sentences is not the same. So, in this field, length of whole samples was set equal to longest sample, and empty space was loaded by zero.

4.1 Feature extraction

Feature extraction is the main important processing, which in fact is the conversions programme of raw signal to feature vector for uses classification. Features are quantities which resulted from sound pre-processing and can be used for sound representation. In proposed method, every sentence was divided to 20 sub words and different features, and related specifications of sub words are extracted and used in speech recognition.

Three groups of separated features are as follows:

LPC is a statistical method by which that we can calculate some different coefficients for forecasting signals with high precision. In this method, we sampling

from each signal and different whole coefficients has been settled by mathematical methods. At first, for extraction of LPC coefficients, we should normalized each signal of sample sentence between 0 and 1, then divide them to 10 blocks and extract suitable LPC coefficient for every block by MATLAB software which we have used length feature of word as sound features.

Frequency features of sound signals are extracted by Fourier transform. So, we divide and separate each sentence to 10 blocks for better extraction of frequency features, then extract their coefficients by Fourier transform.

Pitch and intensity are two features of sound. Sound pitch depends on its frequency. The intensity of sound measured by db, and is magnitude of frequency at air-pressure which is caused by sound waves. After extraction of these features by PRAAT software, we can store them in a file and considered them as network input.

5 Results and Discussion

In this research, we recorded the sound of 14 persons and every individual repeated one word 20 times and this function was done in environment with acoustic walls due to noise omission. Also, according to accomplished researches it was cleared that acoustic features of individuals when saying some letters containing vowel sounds would be distinguished better and it's because of pitch feature that peer better in vowels. We considered 2 individuals as authorized and the other 12 individuals as unauthorized among these 14 persons. Also, we have used 4 sounds of new individuals as unauthorized whose acoustic features were not in training data. Obtained results with ANFIS in different stages of coefficients training could be observed in Table 1, 2 and 3. We changed radius value of ANFIS then compared results in every Table. In these Tables, P1 and P2 show authorized individuals and P3- P14 show unauthorized individuals and P15-P18 are new unauthorized persons who weren't in data training.

According to obtained results in our proposed method, it was clear that represented method producing results towards the others about gathered data. Also, obtained present results have been trained and tested with limited data. We can use more data for better examination and obtain optimum results in this field.

References

1. Anonymous, 2004. Door-access-control System Based on Finger-vein Authentication. Hitachi Review. Available online at http://www.hitachi.com/rev/archive/2004.
2. Yingen Xiong, Speaker Identification System using HMM and Mel Frequency Cepstral Coefficient, Pattern Recognition and Clustering, May 2006
3. Campbell, J.P : Speaker Recognition: a Tutorial. In Proc. IEEE., pp: 1437-1462. 1997.
4. Nalini K.Ratha, Andrew Senior and Ruud M.Bolle, IBM Thomas J.Watson Research Center :Automated biometrics,

5. Wahyudi, Winda Astuti, Syazilawati Mohamed : Intelligent Voice-Based Door Access Control System Using Adaptive-Network-based Fuzzy Inference Systems (ANFIS) for Building Security, Journal of Computer Science 3 (5): 274-280, 2007
6. Jang, J.S.R., C. T. Sun and E. Mizutani. Neuro-Fuzzy and Soft Computing, Prentice Hall, Upper Saddle River, NJ, USA. 1997.
7. Jang, J.S. ANFIS: Adaptive-network-based fuzzy inference system, IEEE Trans. on System, Man, and Cybernetics, vol. 23, pp.665-685, May/June, 1993.

Table 1: Comparison between the LPC results with different radius by ANFIS

	Radius=0.5 ANFIS		Radius=1 ANFIS	
	FAR	FRR	FAR	FRR
P1	-	10	-	10
P2	-	10	-	20
P3	2		4	
P4	0		0	
P5	0		0	
P6	0		0	
P7	0		0	
P8	7		3	
P9	0		0	
P10	-	0	-	0
P11	-	19	-	10
P12	0		0	
P13	0		0	
P14	4		4	
P15	0		5	
P16	0		0	
P17	0		0	
P18	1		4	

Table 2: Comparison between the FFT results with different radius by ANFIS

	Radius=0.5 ANFIS		Radius=1 ANFIS	
	FAR	FRR	FAR	FRR
P1	-	20	-	18
P2	-	40	-	40
P3	4		1	
P4	6		1	
P5	0		0	
P6	4		0	
P7	0		0	
P8	4		0	
P9	4		0	
P10	-	10	-	10
P11	-	0	-	10
P12	4		0	
P13	5		5	
P14	0		0	
P15	4		0	
P16	0		0	
P17	0		6	
P18	4		9	

Table 3: Comparison between the Pitch and intensity results with different radius by ANFIS

	Radius=0.5 ANFIS		Radius=1 ANFIS	
	FAR	FRR	FAR	FRR
P1	-	0	-	0
P2	-	0	-	0
P3	0		0	
P4	12		0	
P5	0		0	
P6	0		0	
P7	0		0	
P8	12		4	
P9	0		0	
P10	-	0	-	0
P11	-	0	-	0
P12	0		0	
P13	2		2	
P14	0		0	
P15	1		0	
P16	0		0	
P17	1		8	
P18	20		0	

Table 4: Comparison between the combination of LPC, FFT, pitch and intensity results from hierarchy ANFIS with different radius by HANFIS

	Radius=0.5 HANFIS		Radius=1 HANFIS	
	FAR	FRR	FAR	FRR
P1	-	0	-	0
P2	-	0	-	0
P3	0		1	
P4	0		0	
P5	0		0	
P6	0		0	
P7	0		0	
P8	0		0	
P9	0		0	
P10	-	0	-	0
P11	-	0	-	0
P12	0		0	
P13	0		0	
P14	0		0	
P15	3		0	
P16	0		0	
P17	4		0	
P18	0		0	