Journal of Theoretical and Applied Information Technology
30th November 2013. Vol. 57 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

LPSO-WNN DENOISING ALGORITHM FOR SPEECH RECOGNITION IN HIGH BACKGROUND NOISE

LONGFU ZHOU[1,2], YONGHE HU[1,2,3], SHIYI XIAHOU[3], WEI ZHANG[3], CHAOQUN ZHANG[2] ZHENG LI[2], DAPENG HAO[2]

1,The Department of Medical Engineer, The General Hospital of Chengdu
No.270, Rongdu Road, Jinniu District, Chengdu 610083, Sichuan, P.R.China;
2,Center for Disease Control and Prevention of Chengdu Command
No.118,Daguan Road,Kunming 650032,China ;
3,School of Automation Engineering, University of Electronic Science and Technology of China
No.2006, Xiyuan Road, High-technology District, Chengdu 611731, Sichuan, P.R.China
Phone: +86-28-86570179, Email: zhoulf_1977@163.com

ABSTRACT

This paper introduces an intelligent evaluation method based on improved PSO-WNN (partiele swarm optimization-wavelet neural network) for speech denoising in high background noise. Firstly, by using Lyapunov stability theory, convergence conditions of a single particle is discussed and based on the result, a new strategy is introduced to improve the performance of the PSO algorithm. Then, the improved PSO algorithm is used to optimize the parameter of the WNN and the LPSO-WNN is introduced. Finally, the trained LPSO-WNN is used to identify and recognition the speech signal in high background noise. Experimental results show that the new method is high efficient and practicable for filtering the high background noise and recognition the speech signal.

Keywords: *LPSO, WNN, Speech Recognition, High Background Noise, Denoise*

1. INTRODUCTION

Automatic speech recognition (ASR) is one of the leading technologies serving as a man-machine interface for real-world applications. Speech recognition system is very sensitive to background noise, when the background noise is large, the speech recognition rate will drop. How to recognize speech signal and improve the recognition rate in high background noise is one of the urgent problems in speech recognition field.

There has been considerable research devoted to the modeling of the functional roles of peripheral auditory systems. Doh-Suk Kim[1] suggested a ZCPA model which is composed of cochlear bandpass filters and a nonlinear stage at the output of each bandpass filter. Qiang Huo[2] introduce a decision strategy called Bayesian predictive classification for robust speech recognition where unknown mismatch between training and testing conditions exists. Reference [3] proposes a noisy speech recognition scheme which generalizes the two-dimensional (2-D) cestrum (TDC) to modified TDC (M_TDC) for use in noisy environments.

In the paper, based on the ingle particle convergence conditions, a new improved PSO algorithm is introduced and used in the training of parameter of the WNN. Then , the trained LPSO-WNN is used to identify and recognition the speech signal in high background noise.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

2. LPSO-WNN

2.1 The Structure Of Wavelet Neural Network

In wavelet neural network, Sigmoid function is replaced by wavelet basis as activation function in cell number. Through linear superposition of wavelet basis, signal representation is achieved. $f(t)$ is fitted by wavelet basis h(a,b, t) as follow.

$$\overline{f}(t) = \sum_{k=1}^{K} \omega_k h\left[\frac{t - b_k}{a_k}\right]$$

Where, $\overline{f}$ is the estimated value of original signal f; ω_k ($1 \le k \le K$) is the weight from hidden layer to output layer; a_k is the shrinkable factor of wavelet basis in hidden layer; b_k($1 \le k \le K$) is the displacement factor of wavelet basis in hidden layer; K is the number of wavelet basis.

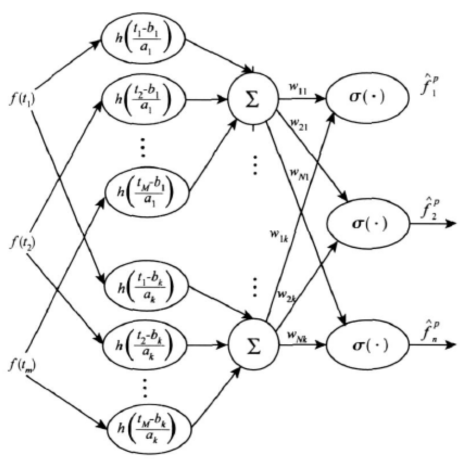


Fig. 1 Structure Of Classified Wavelet Neural Network

In the paper, the classified wavelet neural network in Fig.1 is used. The network is built by two parts: in order to extract characteristic of the signal, signal is decomposed into different channels by wavelet telescopic translation; training the network using signal characteristics.

Set the total number of samples as P, output node of network as N. the relationship of sample and output node is defined as:

$$\hat{f}_n^p = \sigma\left[\sum_{k=1}^{K} \omega_{nk} \sum_{m=1}^{M} f^p(t_m)\, h\left[\frac{t_m - b_k}{a_k}\right]\right]$$

Where, $\sigma(x) = 1/[1 + \exp(-x)]$

The parameters ω_k ,a_k and b_k are optimized by function $E = \frac{1}{2}\sum_{p=1}^{P}\sum_{n=1}^{N}\left(f_n^p - \hat{f}_n^p\right)^2$

2.2 LPSO

In 1995, the population-based algorithm PSO is initiated by Kennedy[4]. The basic PSO is similar in spirit to birds migrating in a flock toward some destination, where the intelligence and efficiency lies in the cooperation of an entire flock. The algorithm updates the entire swarm at each time step by updating the velocity and position

of each particle in every dimension as followed:

$$v_{id}(k+1) = v_{id}(k) + \phi_1(k) * (p_{id}(k) - x_{id}(k)) + \phi_2(k) * p_{gd}(k) - x_{id}(k) \quad (1)$$

$$x_{id}(k+1) = x_{id}(k) + v_{id}(k) \quad (2)$$

By substitution, Eqs.(1) and (2) yield the following equation, defining the trajectory of a particle in a PSO system:

$$x(k+2) - [1 + \omega - \phi(k+1)]\, x(k+1) + \omega x(k) = \phi 1(k+1) * pb(k+1) + \phi 2(k+1) * gb(k+1) \quad (3)$$

where $\phi(k) = \phi 1(k) + \phi 2(k)$.

Equ(3) is a second order difference equation with variable coefficients. Next, by using Lyapunov stability theory (Lyapunov's second method) the stability analysis for Equ(3) is undertaken. Then, stability conditions of particle's trajectory in PSO as Equ(4) and Equ(5) is obtained[5].

$$0 \le \phi(k) < 2 + 2\omega \quad (4)$$

$$|\phi(k+1) - \phi(k)| < 1/10(1 - \omega)\, \phi(k+1)\, (2 + 2\omega - \phi(k+1)) \quad (5)$$

As biologist Couzin had revealed that the larger the group the smaller the proportion of informed individuals needed to guide the group, and that only a very small proportion of informed individuals was required to achieve great accuracy[6].

Journal of Theoretical and Applied Information Technology
30th November 2013. Vol. 57 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Following the concept above, a novel method “particle swarm optimizer with leadership (LPSO)” is introduced to find the global optimum solution with less iterations[7]. The novel method is roughly carried out in three steps: the leader selection, the exploration search and the exploitation search.

2.3 The Algorithm Steps Of LPSO-WNN

In the paper, the LPSO introduced above is used to optimize training steps of wavelet neural network. The algorithm steps are as follows:

Step1: A wavelet neural network according to the network structure shown in Fig.2 is built. Morlet is used as the wavelet function, which is a Gauss wave cosine modulation with high resolution in time domain and frequency domain simultaneously.

$$h(t) = \cos(1.75t)\exp(-t^2/2)$$

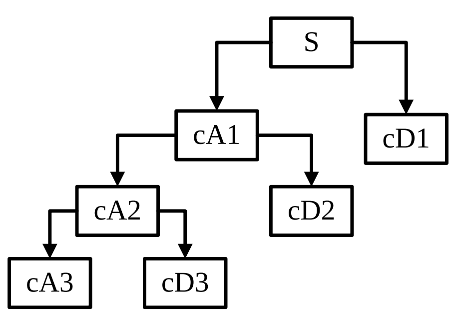


Fig.2 The Model Of Wavelet Decomposition

Step2: Sample selection and pretreatment. In order to ensure the learning fitting precision and reduce the training time, the sample should reflect the various processing of object. Sample pretreatment includes sample data normalization, etc.

Step3: Initialization the wavelet neural network parameters. Firstly, set the cell number in the input layer, hidden layer and output layer; secondly, set wavelet neural network parameters (a1,b1,w1),…, (ak,bk,wk) as position vector of each particle in the algorithm of LPSO.

present(i)
=[w1,w2,…,wk,a1,a2,…,ak,b1,b2,…,bk]
where, k is the number of cell number in the hidden layer of wavelet; i is the number of particle in the algorithm of LPSO.

Step4: definition the fitness function J(d) of particle in LPSO. Sum of the squares of errors between actual output and ideal output after d time iterations is defined as fitness function J(d).

$$J(d) = \frac{1}{2}\sum_{p=1}^{P}\sum_{n=1}^{N}\left[f_n^p - \left(\hat{f}_n^p\right)^d\right]^2,$$

$d = 1,2,\ldots,D$
Where, D is the maximum iteration time.

Step5: searching the optimum solution of the fitness function in Step 4 using LPSO algorithm;

Step6: in the searching process, determine whether the termination condition is reached. If reached, jump to Step 7; if not, to Step 5.

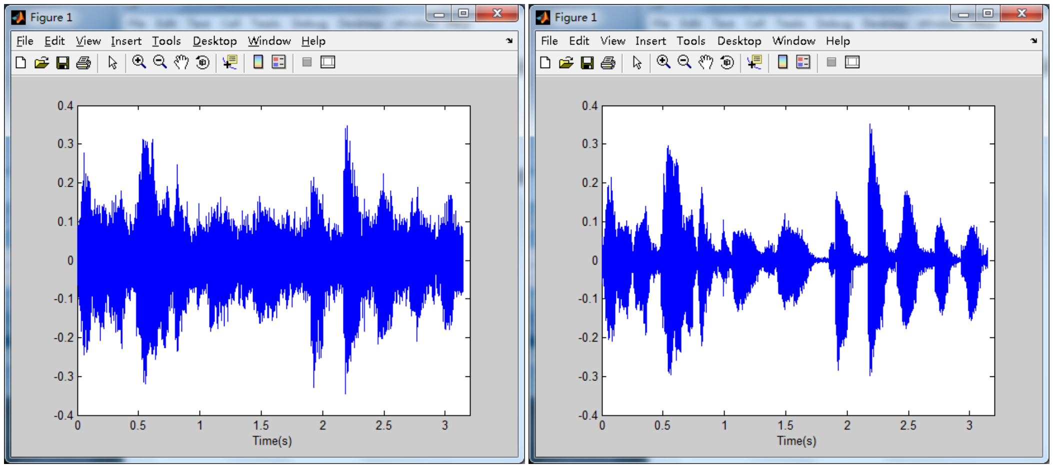
Step7: searching process terminated. The extreme values in Step 4 and corresponding wavelet neural network parameters(a1,b1,w1),…, (ak,bk,wk) are output.

3. SIGNAL PROCESSING AND SIMULATION

Wavelet transform is often used to eliminate the noise. By multi resolution decomposition of Mallat, wavelet transform is realized. Through the low pass filter H1 (z) H0 (z) and high pass filter H1 (z), the signal is decomposed into different frequency range. Then a serial of subband signal is resulted. Because the orthogonal transform has the function of removal of signal correlation and signal energy, signal energy is concentrated to the coefficients of some band through the wavelet transform. So, by set the wavelet coefficients of other bands to zero or give a small weight, the noise can effectively filtered.

Journal of Theoretical and Applied Information Technology
30th November 2013. Vol. 57 No.3

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195



(A) Primitive Speech Signal *(B) LPSO-WNN Denoised Speech Signal*

Fig 3. The Comparison Of Primitive Speech And LPSO-WNN Denoised Speech Signal

In the simulation process, a digital speech database is created with sampling rate of 9.6kHz, 32bits, two channel stereo sound. 10 different people in each section of the audio signal (5 male and 5 female) per 10 times sound is record. In the database, 50 samples are select as training set, 50 as test set.

The experiment was divided into denoising, training and recognition. LPSO-WNN is used to decompose and reconstruct each section of the speech signal. Results show that noise component is removed and the original sound stays clean and undistorted. Figure 3 shows the significant effect on inhibiting the noise by LPSO-WNN.

4. CONCLUSIONS

A feature extraction method motivated by auditory periphery is proposed. The proposed LPSO is based on the theory results of microscopic study on PSO. Next, the new algorithm LPSO is used to train parameter of WNN. Then, the LPSO-WNN is used to identify and recognition the speech signal in high background noise. Experimental comparisons of the developed auditory recognition model with the LPSO-WNN in various types of high noisy environments have demonstrated greatly filtering results, especially in environments corrupted by white Gaussian noise.

ACKNOWLEDGMENT

The authors would like to the support of the key medical science and technology project plan of 12th five-year plan in PLA (BWS11J067) and the major projects in Chengdu military region "Twelfth Five Year" plan of medical research (A12003).

REFERENCES:

[1]. Doh-Suk Kim,Soo-Young Lee,and Rhee M. Kil. Auditory Processing of Speech Signals for Robust Speech Recognition in Real-World Noisy Environments [J].IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, 1999,7(1):55-69.

[2]. Qiang Huo and Chin-Hui Lee. A Bayesian Predictive Classification Approach to Robust Speech Recognition[J]. IEEE

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, 2000, 8(2):200-204

[3]. Chin-Teng Lin , Hsi-Wen Nein, and Jiing-Yuan Hwu. GA-Based Noisy Speech Recognition Using Two-Dimensional Cepstrum[J]. IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, 2010,8(6):664-674.

[4]. Kennedy J, Eberhart R.Particle swarm optimization. In: Proceedings of IEEE international conference on neural networks[C], Perth, 1995:1942–1948

[5]. Longfu Zhou, Yibing Shi, Yanjun Li. Parameter Selection, Analysis and Evaluation of an Improved Particle Swarm Optimizer with Leadership[J]. Artificial Intelligence Review, 2010 (34):343–367.

[6]. Couzin ID, Krause J, Franks NR, Levin SA. Effective leadership and decision making in animal groups on the move [J]. Nature.2005, 433:513–516

[7]. ZhangWei, ShiYibing, Zhou Longfu, Lu Tao. Wavelet neural network classifier based on improved PSO algorithm [J]. Chinese Journal of Scientific Instrument. 2010, 31(10):2203-2209.