

Research on the Voiceprint Recognition Based on Bp-Ga Algorithm

Yanfeng Kong^{1,2}, Lianxing Jia¹, Jiang Zhang¹

¹National Defense Information Institute, Hubei Wuhan, China

²92815 army branch of weapon, Xiang Shan, China

^a13407160738@163.com, ^bjialianxing@163.com, ^cliondangcing@qq.com

Keywords: voiceprint recognition, artificial intelligence, BP-GA, wavelet transform.

Abstract: In order to improve the performance of voiceprint recognition system, the paper proposes to use BP neural network-genetic algorithm (BP-GA) in voice recognition. The algorithm can overcome the problem that the traditional multi-layer artificial neural network is easy to fall into local minima when it is trained by genetic algorithm. The experimental results show that the BP-GA algorithm has the advantages of faster recognition rate, higher recognition rate and lower error rate, automatic error correction and robustness for different speakers, than the traditional recognition algorithms (LPCC, MFCC, etc.).

1. Introduction

With the rapid developing of network technology, information security is becoming more and more important. Conventional password authentication has revealed some shortcomings in using of information network. However, the technology of biometrics has become increasingly mature and has shown its superiority in practical application. Among them, voiceprint recognition is a new recognition technology developed in recent years. Compared with other biometrics, voiceprint recognition has many advantages such as simple, accurate, economical and non-contact identification.

BP-GA algorithm is proposed to identify voiceprint in the paper. Compared with the traditional recognition algorithm, it has the advantages of fast recognition, high recognition rate, low error rate, automatic error correction and robustness for different speakers.

2. Vector processing for voice wavelet transform

As BP-GA recognition algorithm identifies digital variables, the voice signal must be digitally vector-preprocessed. Wavelet analysis is a new concept of time-frequency analysis method, with a variable resolution, reflecting non-stationary transient changes accurately, anti-noise interference ability and other advantages. It can better reflect the dynamic information of the voice signal, and has the advantages of simple implementation and low computational complexity. And can not only fully reflects the auditory characteristics of the human ear but also accurately reflects the dynamic characteristics of the voice signal, so as to improve the final recognition rate of the voiceprint.

For any one-dimensional voice signal, noisecontinuous wavelet transform (CWT) is defined as the inner product of signal and wavelet basis function, that is,

$$W_F(a, b) = (f, \psi_{a,b}) = \left| a \right|^{\frac{1}{2}} \int_R f(t) \psi\left(\frac{t-b}{a}\right) dt \quad (1)$$

Similar to the short-time Fourier transform, the original signal can be recovered from the wavelet transform of the known voice signal, which is called wavelet inverse transform reconstruction. The inversion formula can be expressed as:

$$f(t) = C_{\psi}^{-1} \iint R^2 W_F(a, b) \psi_{a,b}(t) \frac{da db}{a^2} \quad (2)$$

It is assumed that a one-dimensional voice signal containing noises can be expressed as follows:

$$s(i)=f(i)+e(i), \quad i=0,1,2,\dots,n-1 \quad (3)$$

Where $f(i)$ is the true voice ramp signal, $e(i)$ is a Gaussian white noise or other noise signal, $s(i)$ contains the noise signal and the useful low frequency signal to be extracted. The purpose of signal extraction is to extract the useful from the signal with noise so as to recover the true slowly changing signal $f(i)$ in $s(i)$. In practical engineering, the voice signal usually presents as a relatively stable signal, while the noise signal usually appears as a higher frequency signal. Therefore, we choose a wavelet basis and determine a wavelet decomposition level, and then decompose the signal s into N levels. Considering the actual conditions and the amount of computation, we use $n = 3$. As shown in Figure 1.

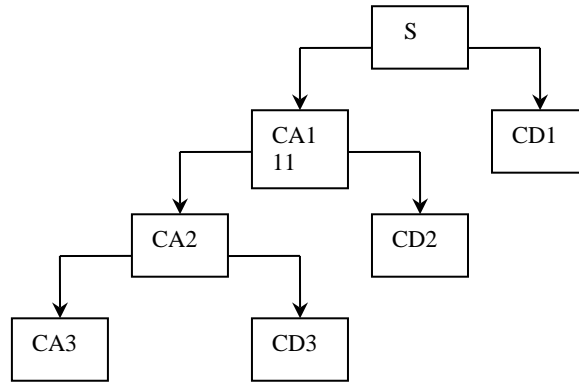


Figure 1. Wavelet signal decomposition diagram

The extracted and decomposed $CD1$, $CD2$, $CD3$ and $CA3$ nodes are denoted $P1$, $P2$, $P3$ and $P4$, respectively. The MFCC parameters are calculated for $P1$, $P2$, $P3$ and $P4$, and the parameter vectors are recorded as $Q1$, $Q2$, $Q3$ and $Q4$.

$$\left\{ \begin{array}{l} (CD1) : P1 \xrightarrow{mfcc} Q1 \\ (CD2) : P2 \xrightarrow{mfcc} Q2 \\ (CD3) : P3 \xrightarrow{mfcc} Q3 \\ (CA3) : P4 \xrightarrow{mfcc} Q4 \end{array} \right. \quad (4)$$

Combining $Q1$, $Q2$, $Q3$, $Q4$ into new vectors $[Q1, Q2, Q3, Q4]$, which is the training voice recognition feature vector used in BP-GA algorithm.

3. Bp-Ga Algorithm

BP-GA algorithm is a new identification method in recent years, with high efficiency, strong recognition ability [3]. The genetic algorithm neither depend on the gradient information, nor require the objective function to be continuous, and may not even need the expression of the objective function. As the combination of artificial neural network and genetic algorithm not only solves the problem of low efficiency and long time consuming in BP neural network, but also plays a global solver of genetic algorithm, so it is an effective and feasible method for identifying. The specific realization of the process shown in Figure 2.

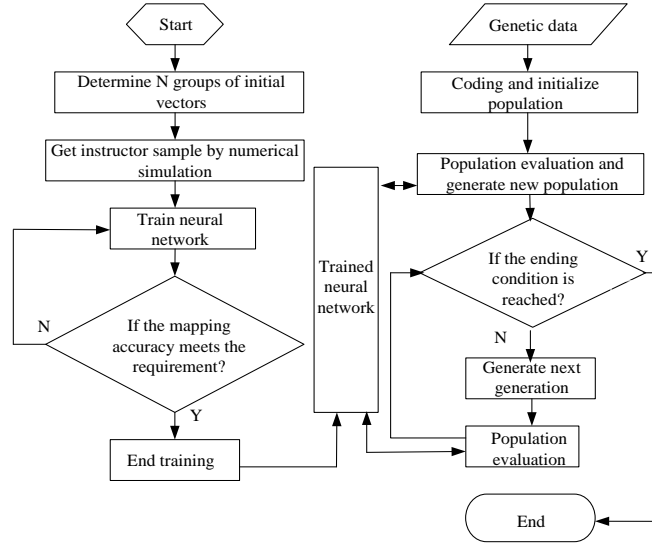


Figure 2 BP-GA algorithm flow chart

3.1 BP neural network construction for voiceprint recognition

BP multi-layer forward-forward neural networks divide the network into several layers, and the layers are arranged in sequence. The neurons in layer i only receive the signal given by the neurons in layer $(i-1)$. Neurons in each layer have no feedback. When a vector x is input to a forward network, a vector y is output after passing through the network. Therefore, the forward neural network can be regarded as an inverter that completes the mapping from x to y . In specific applications, this paper uses a 3-layer BP network. The first layer of the input vector should be adjusted according to the actual situation. The first layer is the normalization layer, the input vector is $[Q1, Q2, Q3, Q4]$. The second layer is the BP network input layer, corresponding to take five nodes. The third layer is the output layer, take a node, the output characteristic function take s-type function, the output value is the credibility of the voiceprint identification, where the continuous number $(0,1)$ interval. ω_{1j} is the input layer weight ($i=1, 2, 3; j=1, 2, 3$), The input layer node is $o1_i$ ($i=1, 2, 3$), ω_{2j} is the weight of the middle layer ($j=1, 2, 3$), $o2_j$ is the middle node ($j=1, 2, 3$), θ_{2j} is the weight of the second middle layer ($j=1, 2, 3$), θ_3 is the output layer weight, ω_0 is the credible value of voiceprint. The mathematical expression of each node is:

$$o2_j = \frac{2}{1 + e^{-u2_j}} - 1 \quad (5)$$

$$u2_j = \sum_{i=1}^3 \omega_{1ij} o1_i - \theta_{2j} \quad (j=1, 2, 3) \quad (6)$$

$$\omega = \frac{1}{1 + e^{-u3}} \quad (7)$$

$$u3 = \sum_{j=1}^3 \omega_{2j} o2_j - \theta_3 \quad (j=1, 2, 3) \quad (8)$$

3.2 Design for Genetic Algorithms

The genetic algorithm includes five basic elements: coding, initial population, fitness function, genetic operation, parameters control and termination rules. Coding is a bridge connecting problems and algorithms. In order to facilitate genetic search in a large space and improve the accuracy of the algorithm, floating-point encoding method are taken. Each individual in the initial population consists of a 40-bit binary string, each of which is generated as follows: A random number between $(0, 1)$ is generated, and if this number is greater than 0.5, the bit code is 1, otherwise 0. The current

method is based on the conventional penalty function idea in the optimization method. Since the paper is the error minimum optimization problem, the fitness function can be expressed as following.

$$f = \frac{1}{\sum_{i=1}^m e_i^2} \quad (9)$$

$$e_i = \omega_i - \bar{\omega}_i \quad (i=1, 2, 3, \dots, m) \quad (10)$$

Genetic manipulation includes selection, crossover, mutation, and group updating in 4 parts.

(1) Selection: Use Roulette for individual selection.

(2) Crossover: Choose a uniform crossover method, with mask samples generated in a random manner.

(3) Variation: Random selection of mutation bits, to flip the bit value.

(4) Group update: Uses Pareto algorithm.

In this paper, we use two termination rules. When the difference between the maximum value and the minimum value of the objective function in the group is less than the given accuracy of $1e-4$, the algorithm is considered to have converged to terminate the program. Otherwise, set the maximum generation to 100, and terminate the program when the number of iterations reaches the value.

4. Figures/Captions

In this case, the comparative experiments are been done by using this recognition algorithm and BP neural network respectively, with the eigenvectors of 100 speaker's voice in the voice database. The object output is 1, the counterfeiter is 0, while training with the identified criterion of 0.5. In the genetic algorithm program In MATLAB 7.0 environment, the initial group of genetic algorithm is 100, cross probability is 0.51, and mutation probability is 0.032. In the BP neural the tansig transfer function is used by hidden layer, the satlin transfer function is used by output layer, a single output type is used as output in order to improve training speed. The general computer configuration is: CPU Pentium M, clocked at 1.4 GHz, memory 512 M, Windows XP.

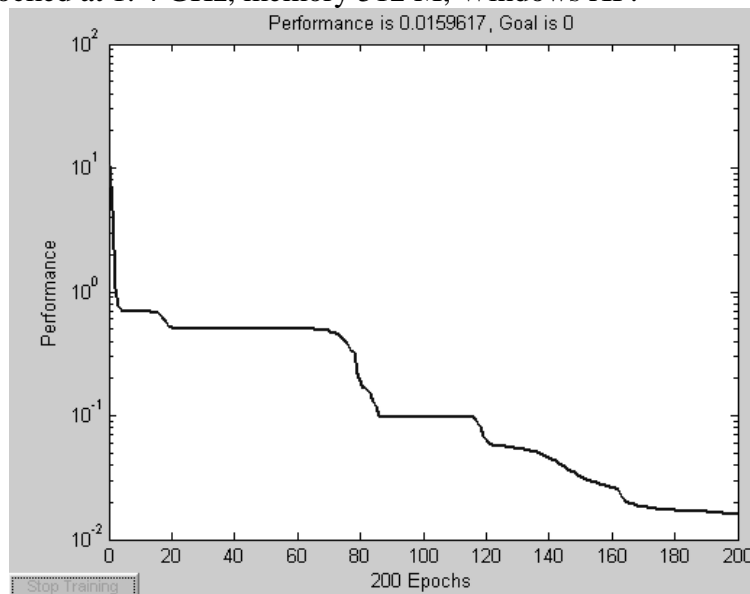


Figure 3 BP neural network training results

Calculated by MATLAB neural network tools, the final relative recognition error rate for the trained BP neural network is about 1.6%, as shown in Fig.3. The specific test results are shown in Table 1. Select the traditional LPCC recognition algorithm as a contrast, the experimental results of

specific tests shown in Table 1.

Table 1: BP-GA pattern voiceprint recognition results contrast with LPCC mode.

Recognition Type	Training Time(s)	Recognition Time(s)	Recognition rate	Error rate
PCC mode	—	12.23	87%	4.4%
BP-GA mode	325.86	9.16	95%	1.58%

5. Conclusion

BP neural network has some intelligence, can gain experience from errors, and improve recognition performance. Genetic algorithm can improve the learning speed, convergence rate of BP neural network, and shorten the time spent on training. Compared with the traditional recognition algorithm (LPCC, MFCC, etc.), the BP-GA algorithm has the advantages of recognition speed, high recognition rate, low error rate, automatic error correction and robustness to different speakers.

References

- [1] Cao Xuelin, Li Jinghua. Voiceprint Recognition Based on Wavelet Packet Extraction, Computer Simulation, 2010, 11(27): 324-327.
- [2] Zhang Qi, et al. Identification of spark plug gap based on wavelet and neural network. Computer Simulation, 2017, 4(34),176-178.
- [3] Koushan K. Automatic hull from optimization towards lower resistance and wash using artificial intelligence. //Proceedings of FAST2003, 7th International Conference on Fast Seat Transportation. Ischia, Italy,2003:120-135.
- [4] Yanfeng Kong, Zhenshan Zhang, Guangtao Cheng. Optimization Design of Elastic Constrained Mechanism in Swashplate Engine Based on Neural Network and Genetic Algorithm. 2010 International Conference on Computer, Mechatronics, Control and Electronic Engineering. Chang Chun: IEEE, 2010.
- [5] Yanfeng Kong, Zhenshan Zhang, Guangtao Cheng. Optimization Design of Elastic Coupling in Swashplate Engine Based on BP-GA. Applied Mechanics and Mechanical Engineering, 2010. 32: 2380-2385.
- [6] WANG Shanhai, JING Xinxing, YANG Haiyan. Research on Speech Recognition of Isolated Words Based on Deep Learning Neural Network. Application Research of Computers, 2015,8(32),2289-2292.
- [7] Deb K, Pratap A, Agarwal S, et al. A fast and elitist multi-objective genetic algorithm: NSGA-II. IEEE Trans. On Evolutionary Computation, 2002, 6(2): 182-197.