Research on Voice Print Recognition Algorithm based on Wavelet-GA-BP Neural Network

Kong Yanfeng¹,² Jia Lianxing¹ and Zhang Jiang²

¹ College of Information and Communication, National University of Defense Technology, Wuhan, Hubei, China
² 92815 Army Branch of Weapon, Xiangshan, Zhejiang, China

13407160738@163.com; simulab@126.com; 103984675@QQ.com

ABSTRACT. Aiming at the fact that common pattern recognition algorithms such as LPCC and MFCC have flaws in voice-print recognition, the paper put forward a new algorithm, which employed wavelet analysis, BP neural network and niche genetic algorithm. Firstly, the algorithm extracted time-domain and frequency-domain characteristic variation of voice signal by wavelet transform, secondly, trained the neural network niche genetic algorithm, which solved the problem of local minimum value caused by normal multi-layers neural network, lastly, took the wavelet transform variations as the training data of the optimization neural network. Simulation experiment was carried out. The results indicate the algorithm in this paper is superior to current recognition algorithms, it has fast recognition velocity, high recognition rate, low fault rate and strong robustness with different voice-makers, besides, it can automatically correct errors.

1. Introduction

As network technology rapidly develops, the information security is more and more important. Traditional password authentication based on personal identification technology has been insufficient, while biometrics technology is increasingly mature and showing its superiority. Among them, voice print recognition is a new technology developed in recent years, it is simple, precise, economical and non-contacted in comparison with other recognition technology [1]. This paper puts forward a new recognition algorithm based on wavelet analysis and BP-GA optimization algorithm, compared with traditional algorithm such as LPCC and MFCC, it has the advantage of fast recognition velocity, high recognition rate, low fault rate, automatic error correcting, and strong robustness with different voice-makers.

2. Wavelet Transform of Voice Signal

For the object of voice print algorithm is digital variation, the pre-processing is needed for voice signal. The LPCC and MFCC algorithm employ FFT frequency-domain analysis method, which can only extract frequency-domain feature, but can not extract time-domain feature, even if add time-window, the time-domain resolution is still inadequate, and the frequency-domain features are easily lost. Wavelet transform is a new time-frequency analysis method, which has the merit of resolution adjustable, strong anti-interference and can reflect signal’s non-stationary and transient feature[2-3]. Besides, it has low computation and is easy to realize, fully reflects not only human ears’s feature but also voice signal’s dynamic feature, and improve the rate of final voice print recognition.
Suppose one division voice signal \( f(t) \in L^1(\mathbb{R}) \), its wavelet transform is defined as the inner product of signal and wavelet base function, that is:

\[
W_{f,a,b}(a,b) = (f, \psi_{a,b}) = \left| a \right|^{-1/2} \int_{\mathbb{R}} f(t) \psi \left( \frac{t-b}{a} \right) dt
\]

(1)

Similar with short-time Fourier transform, the original signal can be re-constructed from its wavelet transform, the formula is:

\[
f(t) = C_v \int_{\mathbb{R}} W_{f,a,b}(a,b) \psi_{a,b}(t) \frac{dadb}{a^2}
\]

(2)

Suppose the signal with noises is:

\[
s(i) = f(i) + e(i), \quad i = 0, 1, 2, \cdots, n - 1
\]

(3)

In eq.3, \( f(i) \) is the real voice signal, \( e(i) \) is gaussian white noise or other noise, \( s(i) \) contains noise signal and useful low frequency signal to be extracted. Voice signals are mainly stable signals in engineering, while noises are mainly high-frequency signals. So we should select a wavelet base and determine the layer of wavelet decomposition. Make 3-layer decomposition of signal, the figure is shown in Fig.1

![Figure 1. Schematic diagram of wavelet signal decomposition](image)

Mark CD1, CD2, CD3, CA3 as d1, d2, d3, a3 and make normalization, mark the normalized parameters as Q1, Q2, Q3, Q4. Construct a new vector \([Q1, Q2, Q3, Q4]\), and this vector is the voice feature vector used for BP-GA algorithm training, the wavelet decomposition result of speech signal is shown in Fig.2.
3. Basic conception of BP-GA

BP-GA is a new optimization algorithm raised recently, it has the characteristics of high efficiency, strong global searching capability. GA is independent of gradient information, and has no requirement of the object function’s continuity, even its definite expression [4]. BP-GA takes the advantage of artificial neural network and genetic algorithm, not only overcomes the low efficiency and long convergence time of GA in structure optimization, but also improves global solving capability of GA, so it is an effective and applicable solution[5-6]. Its working process is shown in Fig.3.

![Figure 3. Flow chart of BP-GA](image)

3.1 Construction of BP neural network voice print recognition

Multi-layer feed forward neural networks divides the network into several layers, which are arrayed in order, the neuron in layer i only receives signals from neuron in layer (i-1), and the neurons in all layers have no feedback. For a forward network, when vector \( x \) is inputted, vector \( y \) is outputted through the network, so the forward neural network can be seen as a transformer of \( x \) to \( y \) mapping[7]. The neural network’s topological structure is shown in Fig.4.

![Figure 4. Topological Structure of BP Neural Network](image)
Three layers BP neural network is employed. The input vector in the first layer can be adjusted according to practice. The first layer is a normalized layer, where three nodes are taken. The input vectors are software threat, hardware threat and management vulnerability. Experts are organized to form an evaluation group to make an evaluation of each factor. The values in \((0,1)\) are supposed as the risk evaluation results, i.e. the better the results are, the lower the risk of information subject’s success is. The second layer is the input layer of BP network, five nodes are taken. The third layer is the output layer, one node is taken, the output characteristic function is s-style function, the output value is information construction risk degree, which are the continual numbers in \((0,1)\).

\[
\omega_{1,j} (i = 1, 2, 3; j = 1, 2, 3) \text{ is the weight of input layer, } \omega_{1,i} (i = 1, 2, 3) \text{ is the node of input layer, } \omega_{2,j} (j = 1, 2, 3) \text{ is the weight of middle layer, } \omega_{3,i} (i = 1, 2, 3) \text{ is the node of middle layer, } \theta_{2,j} (j = 1, 2, 3) \text{ is the weight of the second middle layer, } \theta_{3} \text{ is the weight of output layer, } \omega_{o} \text{ is risk degree.}
\]

The mathematical expressions of the node are:

\[
\begin{align*}
o_{2,j} &= \frac{2}{1 + e^{-u_{2,j}}} - 1 \\
\omega_{2,j} &= \sum_{i=1}^{3} \omega_{1,j} \cdot o_{1,i} - \theta_{2,j} (j = 1, 2, 3) \\
\omega &= \frac{1}{1 + e^{-u_{3}}} \\
\omega_{3} &= \sum_{j=1}^{3} \omega_{2,j} \cdot o_{2,j} - \theta_{3} (j = 1, 2, 3)
\end{align*}
\]  

\[(4) \quad (5)\]

### 3.2 Design of genetic algorithm

Genetic algorithm contains six basic parts: coding, initial population, fitness function, genetic manipulation and parameter control and end rules \([8]\). Coding is the bridge of problem and algorithm. In order to carry out genetic searching in a big space and improve algorithm accuracy, float-encoding is employed. Each individual of the initial population is composed of 40-bit binary strings, the value is determined as follows: generate a random number in \((0,1)\), if it is bigger than 0.5, the value of this bit is 1, or it is 0. Now the most used method is based on the penalty function idea. As the problem talked about in this paper is the least error optimization problem, then the fitness function can be expressed as:

\[
f = \frac{1}{\sum_{i=1}^{M} e_{i}} = \omega_{i} - \omega_{i} (i = 1, 2, \ldots, m)
\]

\[\text{(6)}\]

Genetic manipulation contains five parts, they are selection, crossover, changing step, mutation and population update.

1. Selection: individual is selected by roulette wheel method.
2. Step setting: using scaling method
3. Crossover: in an even crossover way, the mask sample is generated at random.
4. Mutation: mutation bit is selected randomly, then get the bit value reversed.
5. Population update: using pareto solution reserve strategy, the parent pareto solution is taken directly into child generation.

Two terminated rules are adopted here, for one rule, when the difference of the maximum and minimum of the object function is smaller than given precision \(1e^{-6}\), the algorithm is considered convergent, and the program is terminated. For the other rule, suppose the largest genetic generation value is 500, when the iteration time of algorithm reaches 500, the program is terminated. The operation results are shown in Fig.5.
4. Simulation analyses of an example with BP-GA algorithm

Take 200 samples in speech database as feature vector in this example. The object output of real man is 1, and the false man is 0, the criterion for recognition is 0.5. make up MATLAB simulation procedure, the initial population is 500, crossover rate is 0.5, mutation rate is 0.042, the step scale is 0.8. take tansig function as the transfer function in the network’s anonymous layer and satlin function as the transfer function in the output layer, which is single output to improve training speed. The Optimized training results of BP neural network is shown in Fig.6.

From Fig.6, the false rate of recognition is about 1.6% using wavelet-BP-GA algorithm. To test the validity of the algorithm, we made a comparison with MFCC algorithm and LPCC algorithm, the detailed data of the experiment are shown in table.1.

| Recognition pattern | Training time/s | Recognition time/s | Recognition rate | False rate |
|---------------------|-----------------|-------------------|-----------------|------------|
| MFCC                | ——              | 11.51             | 89%             | 5.1%       |
| LPCC                | ——              | 12.23             | 87%             | 4.4%       |
| wavelet-BP-GA       | 325.86          | 9.16              | 95%             | 1.58%      |
5. Conclusions
Wavelet analysis can fully extract the signal’s time-domain and frequency-domain features; BP network can make the recognition system perfect by gaining experience from fault; GA algorithm can improve the learning speed and convergence rate, decrease the network’s false rate, and shorten the voice system’s self-learning time. In short, compared with LPCC and MFCC, the wavelet-BP-GA pattern has the advantage of fast recognition velocity, high recognition rate, low fault rate, automatic error correcting, and strong robustness with different voice-makers.

6. References
[1] Xuelin Cao, Jinghua Li. Wavelet Packet Based Feature Extraction for Voice Recognition[J]. Computer Simulation, 2010, 27(11): 324-327.
[2] Qi Zhang, Shuxia Jiang, Xiangsheng li, et al. Spark Plug Gap Identification Based on Wavelet and Neural Network [J]. Computer Simulation, 2017, 34(4): 176-178.
[3] Koushan K. Automatic hull from optimization towards lower resistance and wash using artificial intelligence[C]. //Proceedings of FAST2003, 7th International Conference on Fast Seat Transportation. Ischia, Italy, 2003: 120-135.
[4] Zhonghua Li, Taishang Zhang. Research of fitness sharing niche genetic algorithms based on extension clustering[J]. Journal of Harbin Institute of Technology, 2016, 48 (5): 178-180.
[5] Yanfeng Kong, Zhenshan Zhang, Guangtao Cheng. Optimization Design of Elastic Constrained Mechanism in Swashplate Engine Based on Neural Network and Genetic Algorithm.[C]. 2010 International Conference on Computer, Mechatronics, Control and Electronic Engineering. Chang Chun: IEEE, 2010.
[6] Yanfeng Kong, Zhenshan Zhang, Guangtao Cheng. Optimization Design of Elastic Coupling in Swashplate Engine Based on BP-GA. [J]. Applied Mechanics And Mechanical Engineering, 2010, 32: 2380-2385.
[7] Shanhai Wang, Xinxing Jing, Haiyang Yang. Study of isolated speech recognition based on deep learning neural networks[J]. Application Research of Computers, 2015, 32 (8): 2289-2292.
[8] Deb K, Pratap A, Agarwal S, et al. A fast and elitist multi-objective genetic algorithm: NSGA-II [J]. IEEE Trans. On Evolutionary Computation, 2002, 6(2): 182-197.