

Design of BP Speaker Recognition System Based on KPCA-MFCC Parameter Optimization

Fengjuan Miao ^{a,*}, Tongri Sun, Bairui Tao ^b, Kaida Liu, Ding Liu, Xiaoxu Lu

College of Communications and Electronics Engineering, Qiqihar University, Qiqihar, Heilongjiang 161006, China

^{a,*} miaofengjuan@163.com, ^b tbr_sir@163.com

Keywords: KPCA, MFCC, BP, Speaker Recognition.

Abstract. The recognition of the speaker through machine learning algorithm has become a hot spot of research. On the basis of speaker recognition based on BP and traditional MFCC characteristic parameters, the feature parameters of MFCC are reduced by KPCA algorithm, and the BP neural network algorithm is used as the back-end recognition model to classify the speaker. The improved algorithm is simulated on the MATLAB platform and compared with the traditional PCA algorithm. The experimental results show that the improved algorithm has a great improvement in recognition efficiency and recognition accuracy and has a good research value.

1. Introduction

The extraction of sound feature parameters has always been the focus of pattern matching research. The quality of feature parameters directly determines the accuracy of recognition [1]. Many scholars have proposed the MFCC to characterize the speech feature information through the study of the auditory mechanism of human ear. This paper will use the MFCC as the characteristic parameter and the BP neural network model as the back-end recognition model for speaker recognition research. The PCA algorithm optimizes the MFCC to reduce the dimension. In order to solve the problem of complex data linear non-separable in the low dimensional space, this paper further introduces the kernel function into the PCA algorithm, uses the KPCA algorithm to reduce the MFCC, and compares it with the traditional PCA algorithm. The recognition rate and robustness of the system are further improved.

2. MFCC Feature Extraction

Each person's voice has its unique characteristics. Figure 1 shows the contrast analysis of the amplitude frequency characteristics and time frequency characteristics of different speaker sounds. It can be seen from the picture that each person's voice is very different in the feature expression. The MFCC is widely used as a token of speech information because it conforms to the auditory awareness of the human ear[2].

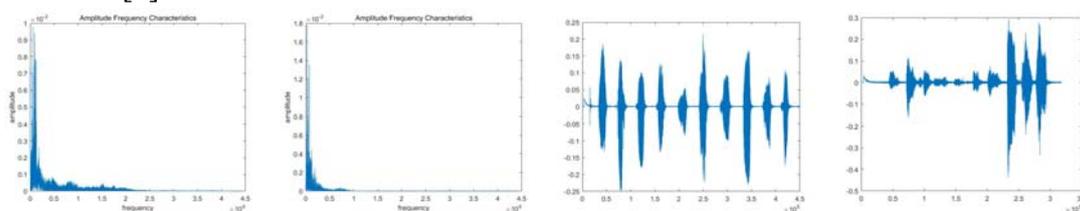


Fig.1 Comparison of time-frequency and amplitude frequency characteristics of different speaker voices

The quality of the speaker's speech feature parameters first determines the performance of the speaker recognition system, so it is very important to extract the MFCC feature parameters of the speech samples accurately. The main steps of extracting the MFCC feature parameters can be divided into several main steps. The flow diagram of the MFCC feature parameter extraction is shown in Figure 2.

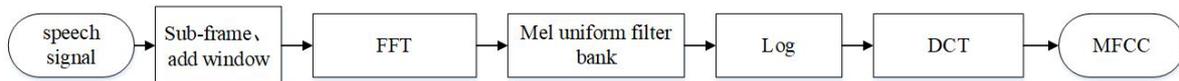


Fig.2 Flow chart of MFCC characteristic parameter extraction

3. Kernel Principal Component Analysis

The core of the PCA algorithm is to calculate the variance in each direction of the data vector separately, take the direction of the maximum variance as the main feature, and make the data in each orthogonal direction unrelated [3]. For M n-dimensional data, we first divide it into columns of n row m column matrix, and the main process of PCA algorithm dimension reduction is as follows:

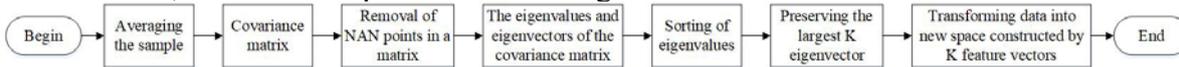


Fig.3 Main flow reduction process of PCA

KPCA uses nonlinear transformation to map the input data from low dimensional space to high dimensional space, so that the nonlinear problem is transformed into a linear problem. Then PCA algorithm is used to extract the main component in high dimensional space, and the purpose of reducing the dimension of the data is achieved on the basis of maintaining the data information [4]. In this paper, the RBF kernel function is applied to the PCA algorithm.

4. BP Neural Network

BP neural network is the error back propagation neural network, its network structure is a multi-level forward feedback network structure, and a simplest BP neural network should contain at least three hierarchies, that is, the input layer, the hidden layer, and the output layer. At each level, there are a lot of small neuron nodes that are used to do some data processing. The number of these small neuron nodes is usually defined by ourselves [5]. The general process of BP neural network for self-learning training is shown in Figure 4.

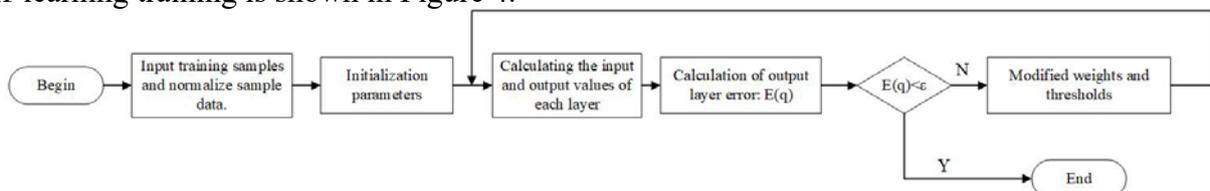


Fig.4 BP neural network algorithm flow

5. Algorithm Simulation

In this study, the data used in this study was recorded by a mobile phone with a format of wav. The length of each speech was 8 to 10 seconds. A total of 17 students were collected as test data. 30 to 40 voice samples were recorded by each student. The overall test sample contained 521 test data.

5.1 MFCC Parameter Extraction and Feature Reduction

When the MFCC parameters are extracted from the test sample, 256 sampling points are taken as one frame, the frame shift is 100, the signal is filtered through 256 Hamming windows, the Mel filter bank is composed of 20 trigonometric filters, the coefficients of second to thirteenth after DCT are taken as the MFCC coefficients. After the MFCC characteristic parameters are obtained, the data are further reduced to 44 dimensions by the PCA and KPCA algorithm. Figure 5 shows the sample data to complete the feature extraction and some data after the dimensionality reduction

optimization.

200x44 double

| | | | | | | | | | | | | | | | |
|---|---------|---------|---------|---------|------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| 1 | -3.3142 | 2.9403 | -1.0908 | 0.4543 | -0.0330 | 0.6585 | 0.2633 | 0.1908 | -0.7626 | 1.8637 | 0.8999 | 0.7943 | -1.2768 | 0.6721 | 0.3501 |
| 2 | -3.2392 | -1.7173 | -0.7221 | 0.5383 | -1.4372 | -0.9856 | 0.2953 | 0.8647 | -0.3875 | -0.5830 | -0.3564 | -0.1530 | 0.0801 | -0.4879 | 0.2894 |
| 3 | 2.0462 | 0.4048 | 1.0845 | 0.6212 | -0.9436 | 0.1291 | -0.6307 | -0.1544 | -0.1552 | 0.3993 | -0.2668 | -0.2452 | -0.8797 | 0.2093 | |
| 4 | -0.6426 | -1.0476 | 3.1848 | -0.5992 | -0.1009 | -0.5777 | -0.3721 | -0.2595 | -0.6192 | 0.0837 | 0.4320 | -0.4674 | 0.1504 | -0.0743 | 0.6617 |
| 5 | -1.2696 | 0.0712 | -1.9253 | -0.5632 | -0.6997 | -0.4075 | 0.7546 | -1.0606 | 0.3620 | 0.5810 | -0.1092 | -1.0292 | -0.2844 | 0.2051 | 0.4177 |
| 6 | 1.9349 | 0.7178 | 0.1833 | 0.5619 | -2.5072... | -0.4652 | 0.1409 | 0.3849 | 0.0176 | -0.0992 | 0.1277 | 0.2459 | 0.7307 | -0.0056 | -0.6965 |
| 7 | 1.5490 | 0.3234 | -0.3770 | -0.5642 | -0.5406 | 0.2396 | -0.0254 | -0.0277 | -0.1002 | 0.2221 | 0.0875 | 0.0765 | 0.9178 | -0.1771 | -0.6847 |
| 8 | -1.0838 | -0.4017 | 0.2580 | -1.8390 | -0.5089 | 1.8449 | 0.1180 | 1.3306 | 0.4315 | 0.9757 | -0.6199 | 0.0379 | -0.7667 | 0.0731 | -1.2212 |

Fig.5 Partial MFCC characteristic parameters

5.2 Construction of BP Speaker Recognition Model

First, the BP neural network framework should be established and the BP neural network model should be trained. The number of training sets is 15000, the error threshold is 0.01. The sample data is randomly scattered and selected as the model training data. First, the sample data should be normalized, then the model is trained by the test sample data, and the remaining 321 sample data are used as test data to test the performance of the training model. The figure 6 shows the result statistics of the speaker recognition using MFCC, PCA-MFCC and KPCA-MFCC as the sample characteristic parameters respectively.



Fig.6 BP speaker system based on the recognition rate of each parameter

Five tests were carried out on the basis of MFCC, PCA-MFCC and KPCA-MFCC characteristic parameters respectively, and the rate and time of each test were recorded. The results were shown as shown in Table 1.

Tab.1 Performance comparison of speaker recognition system based on BP with different characteristic parameters

| | 1 | | 2 | | 3 | | 4 | | 5 | | mean value | |
|-----------|----------|----------------|----------|----------------|----------|----------------|----------|----------------|----------|----------------|------------|----------------|
| | accuracy | time consuming | accuracy | time consuming |
| MFCC | 91.60% | 1.12s | 90.30% | 3.72s | 82.55% | 3.36s | 90.97% | 4.71s | 86.29% | 3.45s | 88.34% | 3.27s |
| PCA-MFCC | 92.80% | 4.61s | 88.20% | 1.22s | 88.70% | 2.05s | 93.10% | 1.28s | 92.10% | 2.5s | 90.98% | 2.33s |
| KPCA-MFCC | 91.90% | 2.17s | 91.59% | 2.09s | 92.21% | 1.11s | 89.10% | 2.12s | 91.28% | 2.14s | 91.22% | 1.93s |

6. Conclusion

On the basis of the conventional BP-MFCC speaker recognition system, the kernel principal component analysis algorithm is used to reduce the dimension of the MFCC feature parameters. The parameters of the optimized parameters are simulated by the MATLAB platform, and the performance is compared with the traditional principal component analysis algorithm. The results show that the recognition accuracy of the speaker recognition system based on the optimized parameters and the efficiency of the test are greatly improved, and it has good research and application value.

Acknowledgments

This work was jointly supported by the Postdoctoral Scientific Research Developmental Fund of Heilongjiang Province (Grant No. LBH-Q15142, LBH-Q14157), Science and Technology Project of Qiqihar (Grant No. GYGG-201409, GYGG-201619), Higher School Science and Technology Achievements Industrialization Pre-Research and Development Foundation of Heilongjiang Province (Grant No. 1254CGZH04), University Nursing Program for Young Scholars with Creative

Talents in Heilongjiang Province (Grant No.UNPYSCT-2016087), Heilongjiang Education Science "Twelfth Five-Year" Plan for the Record (GBGH27),The Research Project on the Reform of Higher Education in Heilongjiang(SJGY20170384), Educational Science Research Project of Qiqihar University(2016090),Scientific Research Foundation for the Returned Overseas Chinese Scholars in Heilongjiang Province, and this work was supported by the Supercomputing Center of Qiqihar University.

References

- [1]. Chen Chen Huang, Wei Gong, Wen Long Fu, Dong Yu Feng. A Research of Speaker Recognition Based on VQ and MFCC[J]. *Applied Mechanics and Materials*,2014,3468(644).
- [2]. Tushar Ratanpara, Narendra Patel. Singer Identification Using MFCC and LPC Coefficients from Indian Video Songs[M]. Springer International Publishing:2015-06-15.
- [3]. Thouraya Hamdi, Adel Ghith, Faten Fayala. A principal component analysis (PCA) method for predicting the correlation between some fabric parameters and the drape[J]. *Autex Research Journal*,2014,14(1).
- [4]. Houying Wang, Weiping Hu. Optimization of Pathological Voice Feature Based on KPCA and SVM[M]. Springer International Publishing:2014-06-15.
- [5]. Xu Chen, Jun Tang. Application of Improved BP Neural Network in the Frequency Identification of Piano Tone[M]. Springer International Publishing:2014-06-15.