Speaker Recognition Based on KPCA and KFCM

Wang Jian
School of Information
Central University of Finance and Economics
Beijing, China
wanderingful@126.com

Zhang Yuanyuan
School of Information
Central University of Finance and Economics
Beijing, China
wanderingful@126.com

Abstract—Speaker recognition system can identify a certain person using speech analysis. Recent advances in speech processing techniques improve the recognition rate. In this paper, an efficient speaker recognition system is proposed. Firstly, a KPCA-based feature selection approach is adopted to get the efficiently reduced dimension of feature vectors and improve clustering performance. Secondly, it has been known that the KFCM has a good superiority in clustering Non-linear and asymmetric samples and it can alleviate the negative influence of the noise and outliers. Thus KFCM clustering algorithm is applied on the selected feature samples to give out a series of clustering centers in feature space, which doubtless can represent the training set in a sense. An analysis is also provided by performing different experiments on the methods that influence the recognition rate. The experiment result shows that the proposed method can resolve the reduce the recognition error rate effectively.

Keywords-Speaker Recognition, KPCA, KFCM,VQ,MCS

I. INTRODUCTION

Speaker recognition system, which includes speaker identification and speaker verification, has many real-world applications, e.g. bank security, telephone transaction, information retrieval, computer login and so on[1].

Feature parameters, obtained using specific signal processing techniques, are the basis of determining the speaker identity, e.g. LPCC(Linear prediction cepstrum coefficient) based on channel, MFCC(Mel frequency cepstral coefficients) based on auditory characteristics , and variations of these methods[2]. The use of feature parameter vector has proved successful for this speaker recognition system application, but if the parameter family is low dimensional, the computational complexity could be reduced and any given member might be represented by a small number of parameters. PCA (Principal Component Analysis) is adopted and the eigenvectors are ordered by the magnitude of their contribution to the variation between the reference speakers [3]. After extracted from training sample, feature vectors are trained to learn and then recognize each speaker's feature parameter characteristic. Speaker clustering has been used to perform the pattern recognition task [4].

An efficient speaker recognition system is proposed in this paper. Feature vectors from speech are extracted by using KPCA and VQ technique is implemented through KFCM algorithm.

II. KERNEL PCA ALGORITHM

By the introduction of nonlinear transformation, KPCA maps the sample vector to the high-dimensional feature space from the input space. Then, PCA is used for feature extraction [5-6]. Given sample data $X_k(k=1,2,\cdots M), X_k \subseteq R^n$, introduce nonlinear function Φ, and realize transformation of $R^n \rightarrow R^l$、 $X_k \rightarrow \Phi(X_k)$; first, suppose in the feature space R^l

$$\sum_{k=1}^{M} \Phi(X_k) = 0 \qquad (1)$$

Then, the covariance matrix in the feature space is shown as follows:

$$\overline{C} = \frac{1}{M} \sum_{j=1}^{M} \Phi(X_j)\Phi(X_j)^T \qquad (2)$$

PCA in R^l is worked out by the solution to Formula (3).

$$\overline{\lambda}w = \overline{C}w \qquad (3)$$

Obviously, as

$$w = \sum_{i=1}^{M} \alpha_i \Phi(X_i) \qquad (4)$$

$$\overline{\lambda}(\Phi(X_k) \cdot w) = \Phi(X_k) \cdot \overline{C}w \qquad (5)$$

The following formula can be obtained according to formulas (2) − (5):

$$\overline{\lambda}\sum_{i=1}^{M} Q_i(\Phi(X_k) \cdot \Phi(X_i)) = \frac{1}{M}$$

$$\sum_{i=1}^{M} \alpha_i(\Phi(X_k) \cdot \sum_{j=1}^{M} \Phi(X_j))(\Phi(X_j) \cdot \Phi(X_i)) \qquad (6)$$

Define $M \times M$ matrix K :

$$K_{ij} \equiv \Phi(X_j) \cdot \Phi(X_j) \qquad (7)$$

Formula (6) can be simplified as follows:

$$M\overline{\lambda}\alpha = K\alpha \qquad (8)$$

Then, the required eigenvalue and eigenvector can be obtained by the solution to Formula (8). The projection of the test sample in the space R^l on the vector W^k is shown as follows:

$$h_k(X) = (W^k \cdot \Phi(X)) = \sum_{i=1}^{M} \alpha_i^k (\Phi(X_i) \cdot \Phi(X)) \qquad (9)$$

$h_k(X)$ is the kth nonlinear pivot component corresponding to Φ .

Make $\overline{\lambda_1} \geq \overline{\lambda_2} \cdots \geq \overline{\lambda_M}$ as the eigenvalue of Matrix K, the number of pivots is generally selected according to the following rule:

$$\left(\sum_{k=1}^{P} \overline{\lambda}_k \bigg/ \sum_{i=1}^{M} \overline{\lambda}_i\right) > E \qquad (10)$$

Namely, the specific value between the first p $\overline{\lambda}_k$ and the sum of total eigenvalue is greater than E. To make KPCA after feature extraction to contain more fault feature information, E is greater than 0.9 in this paper.

In practice, assumed conditions in Formula (1) cannot be met under normal circumstances. At this time, the value of K in Formula (8) is

$$\overline{K}_{ij} = K_{ij} - \frac{1}{M}\sum_{m=1}^{M} K_{mj} - \frac{1}{M}\sum_{n=1}^{M} K_{in} + \frac{1}{M^2}\sum_{m,n=1}^{M} K_{mn} \qquad (11)$$

Please refer to Reference [9] for detailed description on the algorithm. At present, frequently used kernel functions mainly include Gauss radial kernel function, multilayer perception kernel function and p-order polynomial kernel function.

III. KFCM ALGORITHM

A. FCM Algorithm

FCM(Fuzzy C-Means)algorithm is a clustering algorithm based on division, the idea is to make similarity between the objects divided into same cluster is the biggest, while the similarity between different clusters is minimized. FCM algorithm is an improvement of ordinary C-means algorithm, normal C-means algorithm for data classification is rigid, and FCM is a flexible fuzzy partition. FCM divided the n vectors $X_i \ (i=1,2,\cdots n)$ into c fuzzy groups, and find each center of the clusters, making the value function of non-similarity index minimized. The main difference between FCM and HCM is FCM using

fuzzy partition so that each the degree of given data point belonging to various groups can be determined by a partition value between 0 and 1. Adapting the introduction of fuzzy partition, the membership U allows its elements with values between 0 and 1. However, with the normalization requirement, a data set and the total degree of membership equal to 1:

$$\sum_{i=1}^{c} u_{ij} = 1, \forall j = 1,...,n \qquad (12)$$

So the fuzzy clustering problem can be formulated as the following mathematical programming problem:

$$\min J_m(U,V) = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^m d_{ij}^2 \qquad (13)$$

Where m (m>1) is the fuzzy weighting exponent, $d_{ij} = \|c_i - x_j\|$ is the Euclidean distance between the i-th cluster center and the j-th data point. Bezdek gave an iterative algorithm to get the optimal solution of the above mathematical programming problem. The iterative formulae in the algorithm are:

$$c_i = \frac{\sum_{j=1}^{n} u_{ij}^m x_j}{\sum_{j=1}^{n} u_{ij}^m} \qquad (14)$$

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\dfrac{d_{ij}}{d_{kj}}\right)^{2/(m-1)}} \qquad (15)$$

This algorithm is convergent and at the solution point. The objective function attains its local minimum. When running in batch mode, FCM using the following steps to determine the cluster center c_i and membership matrix U:

Step 1: initialize the membership matrix U with random number between 0 and 1, which satisfies the constraints within formula（12）

Step 2: calculate the cluster centers $c_i, i = 1, \cdots, c$ using equation (14).

Step 3: According to equation (13) calculate the value function. If it is less than a determined threshold or its value relative to the last change of function value is less than a threshold, the algorithm stops.

Step 4: using (15) calculate the new U matrix. Return to Step 2.

From above, we can it that the performance of the ECM algorithm depends on the initial cluster centers. Therefore, we can use different fast algorithm to initial cluster centers, even runs FCM multiplied.

B. KFCM Algorithm

Dunn spread the Hard C-Means (HCM) to fuzzy situation based on the fuzzy partition of set which was defined by Ruspini. Bezdek spread it to a more general situation and gave a general description of fuzzy clustering

[7]. The fuzzy clustering problem can be formulated as the following mathematical programming problem [8].

where m (m>1) is the fuzzy weighting exponent, c is the number of clusters to be explored, n is the card of dataset X. Bezdek gave an iterative algorithm to get the optimal solution of the above mathematical programming problem. The iterative formulae in the algorithm are:

This algorithm is convergent and at the solution point (i.e. the partition matrix U and the prototypes V) the objective function attains its local minimum.

IV. VECTOR QUANTIZATION

VQ (Vector quantization) is implemented through LBG algorithm, which required number of highly representative code vectors for each speaker

By using the collection of these code vectors which is called a codebook [9].we can construct a model system to realize speaker recognition.

In this system, each speaker is modeled with a vector quantizer, the training databases will have a codebook for each speaker. In testing phase, the identification is done quantizing the input sentence, the unknown voice after extracting voice feature vectors will be compared with the codebook of each speaker in the training set and distortion will be computed. Based on the minimum distance, system will provide the identity of the speaker [10]. The proposed speaker recognition system based on VQ combined KPCA and KFCM is shown in Fig .1.
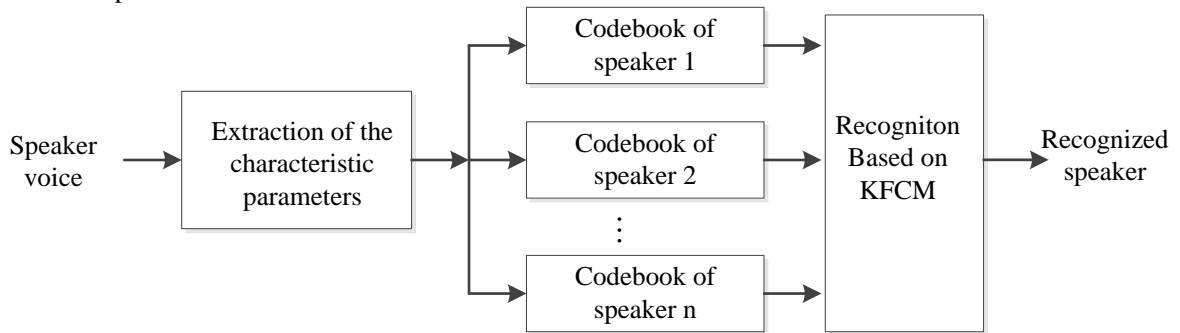

Figure 1. Proposed speaker recognition system

V. EXPERIMENT STRUCTURE AND RESULTS ANALYSIS

Database includes voice samples of Arabic numbers are pronounced by 60 different speakers (30 males and 30females) and recorded in two different sessions with a gap of two weeks is used for the training and testing of the developed system. Voice samples recorded in first session are used as training data. After extracted from a digital voice and selected by KPCA, the cepstral coefficients, is taken as feature vector by vector quantization method to design codebook of each speaker. The second session is used as testing data. According the KFCM formula

described above, each frame of speech is classified by taken each speaker codebook as clustering center.

In the process of calculation, if one of feature vector of a voice clustering to a type of codebook, we can indicate it recorded by the speaker represented by the codebook, and so forth, each feature vector of paragraph speech obtained its codebook category. We statistic all feature vector category to identify the certain type of these feature vectors categories maximum , then conclude that this voice is pronounced by the speaker belongs to the class.

The recognition rate of using VQ method only and using the proposed algorithm is compared, as shown in table1.

TABEL I. ERROR RATE COMPARISON

Error rate(%)	Speaker 1	Speaker 2	Speaker 3	Speaker 4	Speaker5	All speaker
VQ	10.3	12.3	10.6	11.9	10.5	11.3
VQ+KPCA	7.4	6.9	8.0	7.1	7.5	7.2
VQ+KPCA+KFCM	3.3	3.5	3.6	3.7	3.9	3.5

As the proposed algorithm results are shown in Table 2, we can see clearly that based on the combination of KPCA and KFCM, more than 90 percent of the feature vector of a certain speaker can be clustered in ones codebook, which

can correctly determine a certain period of voice by which speaker said. This method is an effective speaker recognition system with performance has been significantly improved compared with using VQ only.

TABEL II. RESULT OF SPEAKER RECOGNITION BASED ON THE PROPOSED METHOD

Clustering rate	Codebook of speaker1	Codebook of speaker2	Codebook of speaker3	Codebook of speaker4	Codebook of speaker5
Voice of speaker1	96.7	1.07	0.86	0.67	0.7
Voice of speaker2	0.99	95.9	0.88	1.3	0.93
Voice of speaker3	0.54	0.98	96.5	0.86	1.12
Voice of speaker4	0.86	0.54	1.1	96.9	0.6
Voice of speaker5	0.56	1.2	1.46	0.98	95.8

VI. Conclusion

Results show that KPCA based feature selection with FCM modeling technique has very good identification accuracy and therefore, it is robust against noise. After analyzing the results of both experiments it is also concluded that sampling frequency of speech and number of vectors in VQ codebook improve the identification accuracy greatly. In future, a Multiple Classifier System (MCS), having more than one classifier, will be designed to further improve the identification accuracy. Hidden Markov Model(HMM) will be used as a classification technique. Finally, a suitable combination technique will be required to reach a consensus by combining the individual opinions of each classifier.

References

[1] K. R. Farrell, R. J. Mammone, and K. T. Assaleh: IEEE Trans. Speech Audio Processing vol. 2 (1994), p. 194–205I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[2] K.M. Kim, S.Y. Kim, J.K. Jeon, et al: IEEE Trans. on Consumer Electronics vol.52-1 (2006), p.200-205.

[3] C. W. Seo, K. Y. Lee and J. Lee: Electronics Letters vol. 37(2000), p. 1486–1488.

[4] M. Padmanabhan, L. Bahl, D. Nahamoo and M. Picheny: IEEE Transactions on Speech and Audio Processing vol. 6(1998), p.71-77.

[5] B.Scholkopf, A Smola, K.R.Muller:Nonlinear component analysis as kernel eigenvalue problem[J]. Neural Computaion,2008,10(5),p.1299-1399.

[6] JH.Hu,SS.Xie,SL,Hou,et al:Kernel principal component analysis and its application to feature extration[J],Jounal of Vibration,Measurement &Diagnosis,2007,27(1):48-52.

[7] Information on http://www.codeproject.com/Articles/59732

[8] X. Gao, WX. Xie: Development of the Fuzzy Clustering Theory and its Application. (Chinese Science Bulletin,1999).

[9] Y. Linde, A. Buzo, R.M. Gray: IEEE Transactions on Communications vol.COM-28(1980), p. 84-95.

[10] T. Matsui, S. Funii: IEEE ICASSP'92, v01.2 (1992), p.157-160.