Discriminative Feature Extraction based on PCA Gaussian Mixture Models

Hossein Marvi
Shahrood University of Technology
Faculty of Electrical Engineering and Robotics
Shahrood
Iran

Abstract: A discriminative feature extraction based on Principal component analysis (PCA) and Gaussian mixture models is presented to increase the discrimative capability of modified two dimentional root cepstrum analysis (MTDRC). The exprimental results show that F-ratio tests indicate better separability of phonemes by using discriminative feature extraction than MTDRC.

Key–Words: Feature extraction,PCA , Gaussian mixture model,MTDRC

1 Introduction

Feature extraction has a great influence on the efficieny of a speech recognition system. In order to improve the performance of speech recognition, it is desired that dissimilar acoustic vectors would be clearly separable from each other in the feature space and corresponding similar acoustic vectors would be close to each other. Thus, an approach to feature extraction include discriminative based feature extraction, whose objective is to make the represention of different classes as different from one another as possible in the new resulting feature space. The novel contribution described in this paper reports on the result of using a discriminative feature extraction based on PCA Gaussian mixture models to increases the discrimative capability of modified two dimention root cepstrum analysis [8] [7] [9].

2 Discriminative feature extraction

The Discriminative feature extraction representation can be achieved by applying a linear transform matrix such as A to original feature space as

$$y = Ax \tag{1}$$

where x are the original feature vector and y is the transformed ones. In the new representation space, the most discriminative components are enhanced and normally this increases the discrimination capability of recognizer. The transform matrix is desired to satisfy the following characteristics:

1. Remove the redundant information and reduce dimensionality of the feature vectors.
2. De-correlated the features to lead a digitalized covariance matrix during the modeling the feature coefficients.
3. Minimize the classification error.

The discriminative feature extraction has been proposed for improving the representation of the speech signal [2] [5] [3] [6]. The element of the transformation a_{ij}, are iteratively computed with the Minimum Classification Error criterion by a gradient descent algorithm in order to minimize a cost function L which represents the classification error. At iteration l, a_{ij} can be calculated by gradient descent of cost function [4]:

$$a_{ij}^{l} = a_{ij}^{l-1} - \eta \frac{\partial L}{\underline{\partial a_{ij}}} \tag{2}$$

where the η represent the convergence coefficient.

3 Gaussian mixture model

A Gaussian mixture model (GMM) is a weighted sum of several multivariate Gaussian densities and is given by:

$$p(x) = \sum_{j=1}^{M_g} p(x|j)P_j \quad (3)$$

where $p(x|j)$ is a Gaussian family densities and given by:

$$p(x|j) = G(x|\mu_j, C_j) \quad (4)$$

with

$$G(x|\mu, C) = (2\pi^{-D/2})|C|^{-1/2} e^{[-1/2(x-\mu)^T C^{-1}(x-\mu)]} \quad (5)$$

In equation (5) μ is a mean vector, C is a covariance matrix, D is the dimension of vector x and M_g is the numbers of components, . The coefficient P_j is a mixing parameter and represents the weight associated with component function, $p(x|j)$. They are chosen such that:

$$\sum_{j=1}^{M_g} P_j = 1 \qquad and \qquad 0 \leq P_j \leq 1 \quad (6)$$

Thus, a Gaussian mixture model presents each class of data as a linear combination of several Gaussian densities in the features space. The complete Gaussian model is parameterised by the mean vectors, covariance matrices and mixture weights from all component densities. These parameters are collectively represented by the notation:

$$\theta = \{\mu_j, C_j, P_j, \} \qquad j = 1, ..., M_g \quad (7)$$

Given the number of components M_g, the GMM parameters, i.e $\theta = \{\mu_j, C_j, P_j, \}$ are computed by using Expectation Maximisation algorithm [1].

4 PCA Gaussian mixture model

Principal component analysis (PCA) is one of the most frequently used techniques to reduce the dimensionality of a data set [10]. A set of observed D-dimensional data vector $X_N = x_1, x_2, ..., x_N$ and $x_i \in R^D$can be mapped to d-dimensional feature space by a linear transformation matrix A as the following:

$$y_i = A^T[x_i - m_{x_i}] \qquad i \in \{1, 2, .., N\} \quad (8)$$

where m_{x_i} is the mean value of vector x_i, $y_i \in R^d$ and $d << D$. The goal of PCA is to find A that minimize the accumulated squared difference between the projection of the data in the new space and the original data as:

$$A = Argmin\Big\{\sum_{i=1}^{D}\{\sum_{i=1}^{d} A_j x_i A_j - x_i^T\}\{\sum_{i=1}^{d} A_j x_i A_j - x_i^T\}^T\Big\} \quad (9)$$

The resulting matrix A is actually the eigenmatrix of the covarince matrix of the original observed data, i.e: $A = [V_1, V_2, ..., V_d]$ and vector V_i is the eigenvector corresponding to the dominant eigenvalue of the observed data covariance matrix. As PCA assumes a single multivariate Gaussian model for the data it application for complex data such as MTDRC features, which different clusters may need different projection direction, is limited. Therefore, A PCA Gaussian mixture model is desirable which provides a better model. A PCA Gaussian mixture model can be obtained by applying Gaussian mixture model to the feature which have been mapped to PCA space [**?**] and is given by :

$$p(y) = \sum_{j=1}^{M_g} p(y|j)P_j \quad (10)$$

where y is computed by eq.(8) and $p(y|j)$ is a Gaussian density function. As the d principal axes of transfomtion matrix in PCA i.e A are orthonormal axes the PCA feature vector y_i should be uncorrelated and its covariance matrix is a diagonal matrix with dominant eigenvalue of the covariance of original features. The condesional density function for PCA space is cut down to the following:

$$p(y|j) = \prod_{i=1}^{d} \frac{1}{(2\pi)^{1/2}\lambda_i^{1/2}} e^{-\frac{y_i^2}{2\lambda_i}} \quad (11)$$

where λ_i , $i \in \{1, .., m\}$ is the ith dominant eigenvalue of the features. The PCA Gaussin mixture parameter is estimated by maximising the likelihood.

5 Experimental results

To compare the discriminative capability of suggested method a generalized F-ratio method is used [11]. It can be used to evaluate the scene discriminative capability of the selected feature. If this factor is large enough, then the corresponding feature has good class separability. This means that the feature vectors which belong to the same class are close to each other while the feature vector from different classes are clearly separable and far from each other. In deriving the F-ratio separability , let the mean feature vector and sample covariance of ith phoneme classes are denoted by M_i and R_i respectively. Assuming each phoneme class has equal probability, the mean vector of all classes is calculated as $M_0 = 1/I \sum_{i=1}^{I} M_i$ where I denote the number of phoneme classes being compared. The within-class scatter matrix S_w and the between class scatter matrix S_b are computed as:

$$S_w = 1/I \sum_{i=1}^{I} R_i \tag{12}$$

and

$$S_b = 1/I \sum_{i=1}^{I} (M_i - M_0)(M_i - M_0)^T \tag{13}$$

The separability criterion is defined as:

$$J_{tr} = tr(S_w^{-1} S_b) \tag{14}$$

The notation *'tr'* is the trace of matrix which is the sum of the diagonal elements of a matrix and can be used to convert the matrix product $(S_w^{-1} S_b)$ to a scalar. Experiments are carried out over the TIMIT data base for English phonemes. The analysis conditions of speech signal are shown in table(1). The MTRDC

Preemphasis factor	0.95
Analysis window	Hamming
Analysis window size	32 ms
Frame rate	12 ms
Feature set I	Original MTDRC features
Feature set II	Discriminative features

Table 1: Analysis condition of speech signal

analysis has been applied to each phoneme and MTDRC matrix is calculated for each phoneme. Then PCA transform matrix A is also computed for MTDRC features. By using the transformation matrix the MTDRC is mapped to the PCA space. A Gaussian mixture is generated for each PCA space of MTDRC feature by using EM algorithm.
Table (2) shows a list of phonems used in the experiments. Fig.1 shows the separability measure ,J, as a function of the dimesion of feature vector for original feature and discriminative feature based on PCA Guassian mixture models for vowels 'IY', 'IH' and 'EH' as an example. Fig.2 shows the separability for consonants 'B','D','G' and 'P'. It can be seen from the figures that the discriminative feature based on PCA Guassian mixture models provides better separation between phonems than original MTDRC features. The change of separability J with dimension of feature vector is very similar for different phoneme classes. From these figures it can be observed that the consonants have bigger separability than vowels. The gearter separation is probably the reason that the recognition accuracy of consonants is normally higher than that of vowels.

Phone	Example
IY	beat
IH	bit
EH	bet
B	bob
D	dad
G	green
P	pop

Table 2: Phonemes used in the experiments

6 Conclusion

It has been shown in this paper that the PCA Gaussian mixture model provides a discriminative feature extraction for speech recognition application. In the new representation space, the most discriminative features are enhanced and this increases the performance of the speech recognition system.

7 Acknowledgement

The author wish to thank the Shahrood University of Technology, for financial support of this work.

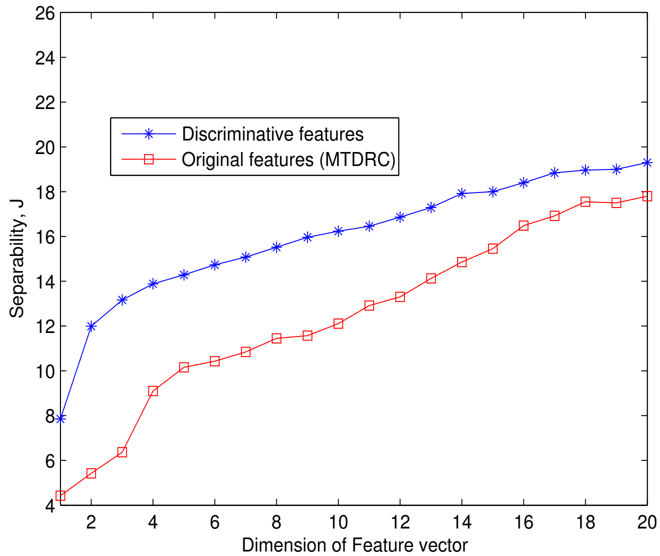

Figure 1: Separability J versus dimension of the feature vector for 3 vowels

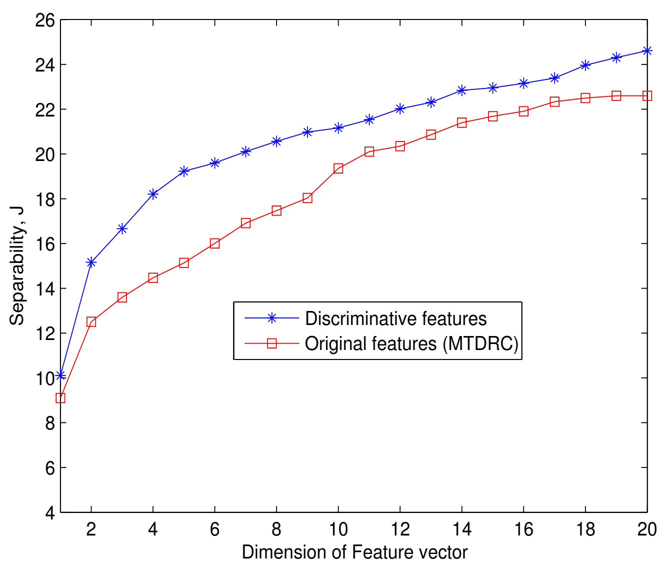

Figure 2: Separability J versus dimension of the feature vector for 4 consonants

References:

[1] N. M. Laired A. P. Dempster and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal statistical society B*, 39:1–38, 1977.

[2] A. Biem and Katagiri. Feature extraction based on minmum classification error generalised probabilistic descent method. In *Proc. of ICASSP*, volume 2, pages 275–278, 1993.

[3] E.L. Bocchieri and J. G. Wilpon. Discriminitve feature selection for speech recognition. *Computer, Speech and Language*, 7:229–246, 1993.

[4] B. Juang and S.Katagiri. Discriminative learning for minimum erro classification. *IEEE trans. on Signal processing*, 40:3043–3054, 1992.

[5] A. De la Torre, A.M.Peinado, A.J. Rubio, V. E. Sanchez, and J.E. Diaz. An application of minimum classification erro to feature space transformation for speech recognition. *Speech Communication*, 20:273–290, 1996.

[6] B. Mak, Y. Tam, and Qi Li. Discriminative auditory based features for robust speech recognition. *IEEE trans. on Speech and audio processing*, Sept. 2003.

[7] H. Marvi and E. Chilton. Acoustic images based on hartley tdrc analysis. In *In MVIP2005 University of Thehran Iran*, 2005.

[8] H. Marvi and E. Chilton. Modified two-dimensional root cepstrum analysis. *IEE, Electronics Letters*, 41(5):285–286, March 2005.

[9] H Marvi, I. Paraskevas, and E. Chilton. Acoustic classification using time-frequency distributions. In *Proceeding of the Institute of Acoustics*, volume 26 Pt.2, pages 612–620, 2004.

[10] Jolliffe I. T. *Principal component analysis*. Springer-Verlag, New York, 1986.

[11] S. Umesh, Leon Cohen, and J. Nelson. Scale transform in speech analysis. *IEEE transaction on speech and Audio processing*, 7(1), 1999.