International Journal of Current Advanced Research
ISSN: O: 2319-6475, ISSN: P: 2319-6505, Impact Factor: SJIF: 5.995
Available Online at www.journalijcar.org
Volume 6; Issue 9; September 2017; Page No. 6298-6301
DOI: http://dx.doi.org/10.24327/ijcar.2017.6301.0913

Research Article

AUTOMATIC SPEECH RECOGNITION SYSTEM OF TAMIL LANGUAGE USING LINEAR DISCRIMINANT ANALYSIS

Sundarapandiyan S* and Shanthi N

Department of Computer Science and Engineering, Kongu Engineering College, Erode

A R T I C L E I N F O

Article History:

Received 12th June, 2017
Received in revised form 3rd July, 2017 Accepted 24th August, 2017
Published online 28th September, 2017

Key words:

Automatic Speech Recognition, Feature Extraction, Linear Discriminant Analysis Principle Component Anaysis

A B S T R A C T

This paper presents a method of dimensionality reduction in Tamil language speech recognition system. The method uses Linear Discriminant analysis (LDA). The intent of using LDA is to reduce the training time of the acoustical model also reduces the size of the feature vector. In General an automatic speech recognition system uses high dimensional acoustic vectors for train the system. LDA converts high dimensional acoustic feature vector to low-dimensional acoustic feature vector. The system uses the standard ELDA Tamil corpus. An automatic speech recognition system using LDA produces better result than the other dimensionality reduction techniques such as Principal Component Analysis (PCA), Multi dimensional Scaling(MDS), Locally-Linear Embedding(LLE) used speech recognition systems.

INTRODUCTION

Automatic Speech Recognition (ASR) is to convert a speech waveform into textual form. The problem can be defined as finding the most probable sequence of words W given the acoustic input O which is computed as:

$$P(W|O) = \frac{P(O\backslash W)P(W)}{P(O)} \qquad (1)$$

Where
$P(O\backslash W)/P(O)$ – Emission probability,
$P(W)$ – Prior Probability

Given an acoustic observation sequence O, classifier finds the sequence W of words which maximizes the probability $P\,(O|W\,).P\,(W\,)$ The quantity $P\,(W\,)$, is the prior probability of the word which is estimated by the language model. $P\,(O|W\,)$ is the observation likelihood, called as acoustic model.

Commonly in speech recognition system uses acoustic vectors in the high-dimensional space can be created by concatenating the MFCC representations of multiple consecutive frames. In this paper we apply the low dimensional acoustic vector for discrimination of different phonemes, which can be used within the speech recognizer and Low dimensional acoustic representations achieved by dimensionality reductions. There are a couple of benefits to using the dimensionality reduction technique. First of all, it can dramatically reduce the word error rate. Second, it also makes the decoder faster since it reduces the dimensionality of the features, and also reduces the size of the acoustic model.

Many dimensionality reduction methods have appeared which can be categorized into linear (e.g. Principal Component Analysis (PCA), Multi-Dimensional Scaling (MDS) and Linear Discriminant Analysis (LDA)) and non-linear (e.g. Locally-Linear Embedding (LLE), ISOMAP, Laplacian Eigen map, Kernel Principal Component Analysis (KPCA)) dimension reduction methods. The differences between these methods lie in their different motivations and objective functions. Linear transform can be expressed as

$$y = \theta^T x \qquad (2)$$

Where
y – A feature vector in the reduced feature space, $y \in Rp$.
X – Original feature vector, $x \in Rn$.
θ – Transformation matrix. The transformation matrix is a n × p matrix. The goal of all feature reduction techniques is to find the optimal value with respect to some optimization criterion. Linear Discriminant Analysis and Principal Component Analysis are two effective methods for dimensionality reduction [1].

Linear Discriminant Analysis easily handles the case where the within-class frequencies are unequal and their performances have been examined on randomly generated test data. This method maximizes the ratio of between-class variance to the within-class variance in any particular data set thereby guaranteeing maximal separability [2]. The use of

Corresponding author:* **Sundarapandiyan S
Department of Computer Science and Engineering, Kongu Engineering College, Erode

Linear Discriminant Analysis for data classification is applied to classification problem in speech recognition. It can be defined as

$$\hat{\theta} = \underset{\theta}{max} \frac{|\theta^T s_b \theta|}{|\theta^T s_w \theta|} \quad (3)$$

Where

s_b – Scatter matrix Between-class
s_w - Scatter matrix within-class
θ - Transformation matrix.

Principal Component Analysis (PCA) is a dimension-reduction tool that can be used to reduce a large set of variables to a small set that contains the information in the large data set. Principal component analysis (PCA) is a mathematical procedure that transforms a number of correlated variables into a (smaller) number of uncorrelated variables called principal components. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible [3]. The objective of PCA is to reduce dimensionality by extracting the smallest number components that account for most of the variation in the original multivariate data. The first component extracted in a principal component analysis is

$$C1 = b11(X1) + b12(X2) + \ldots\ldots. b1p(Xp) \quad (4)$$

Where
$C1$ – The first component extracted
bip – The regression coefficient (or weight) for observed variable p, as used in creating principal component 1
Xp – The subject's score on observed variable p.

Related Work

Many Techniques were implemented to reduce the dimensionality of the acoustic feature vectors. Principal Component Analysis (PCA) to map the variance of the speech material in a database into a low-dimensional space, followed by clustering and a selection technique [4]. A unified algorithmic frame-work for solving many variants of MDS [5]. Locally-Linear Embedding (LLE) was presented including faster optimization when implemented to take advantage of sparse matrix algorithms [6]. The largest improvements in speech recognition could be obtained when the classes for the LDA transform were defined to be subphone units [7].

Problem Statement

PCA does more of feature classification and LDA does data classification. In PCA, the shape and location of the original data sets changes when transformed to a different space whereas LDA doesn't change the location but only tries to provide more class separability. LDA also helps to better understand the distribution of the feature data [8] [9]. The discrimination capability of LLE is less compared to LDA. Compared with unsupervised methods MDS, LDA is prone to over fitting when the training data set is small and the dimension is large .In this experiment we apply this dimensionality reduction technique LDA on Tamil Language.

Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) is a technique used for dimension reduction. Unlike PCA and MDS, however, it is a supervised learning algorithm. Therefore, it considers the class labels (i.e. in addition to knowing a data point's coordinates, the algorithm takes into consideration the point's group membership, whether it be in terms of gene, sample condition, or other factor). When reducing the dimensions, LDA seeks to maximize the separation among the different groups by preserving as much as the class discriminatory information as possible [10] [11]. This means that we must have some measure of the separation between the groups. One possible way is to maximize the distance between the means of the groups when they are projected onto line. As seen in the Figure 1, maximizing the distance between the projection means does not always maximize the distance between the groups because it does not take into account the variability within each group. Instead of this naive approach, LDA maximizes the distance between the projection means while at the same time minimizing the scatter within the group (Figure2).

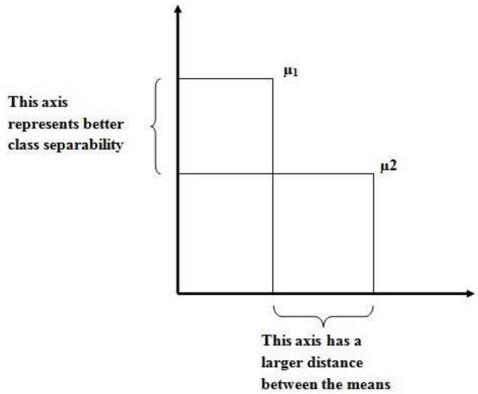

Figure 1 Class Separability and distance between the mean

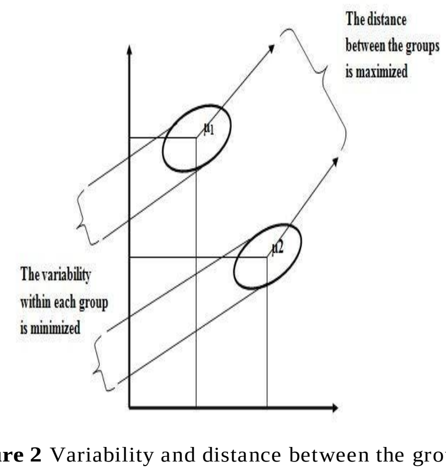

Figure 2 Variability and distance between the groups

System Architecture

Figure 3 shows the architecture of automatic speech recognition for Tamil language using Linear Discriminant Analysis. The training data is given as input to the feature extraction; it produces the high dimensional acoustic MFCC feature vectors. Using LDA this high dimensional acoustic vectors are converted into the low-dimensional acoustic vectors. Low-dimensional acoustic vectors reduces the training time of the acoustical model also reduces the size of the acoustical model [12]. Train the acoustical model by using low-dimensional acoustic vectors we can get posterior probability for each phoneme. Decoder uses the posterior probability and prior probability from the language model recognizes the word from the test data.

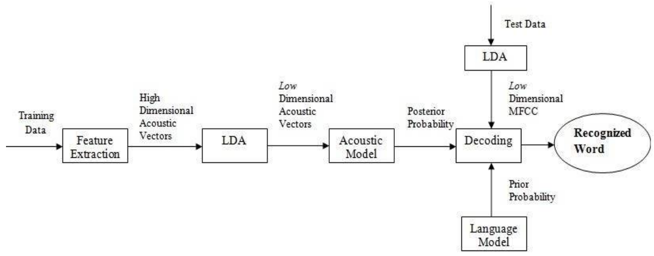

Figure 3 Architecture of automatic speech recognition for Tamil language using Linear Discriminant Analysis

Algorithm of Linear Discriminant Analysis

Step 1 Represent the training data set as matrix consisting of feature in the form of given below

Step 2 Compute the mean of each data set and mean of entire data set. Let $\mu 1, \mu 2, \ldots, \mu n$ be the mean of $set1, set2, \ldots\ldots, setn$ respectively. μ be mean of entire data which is obtained by merging $set1, set2, \ldots, setn$.

$$\mu = p1 \times \mu 1 + p2 \times, \ldots\ldots\ldots, pn \times \mu n \quad (5)$$

Where $p1, p2, \ldots\ldots, pn$ – Apriori probabilities of the classes

Step 3 In LDA, within-class and between-class scatter are used to formulate criteria for class separability. Within-class scatter is the expected covariance of each of the classes. The scatter measures are computed using (6).

$$sw = \sum_j pj \times (covj) \quad (6)$$

$$covj = (xj - \mu j)(xj - \mu j)T \quad (7)$$

Step 4 The between-class scatter is computes using the following equation

$$sb = \sum_j (\mu j - \mu) \times (\mu j - \mu)T \quad (8)$$

Step 5 LDA maximizes the ratio between-class variance to the within-class variance with this definition we can easily formulate the optimization criterion, namely

$$\hat{\theta} = \underset{\theta}{max} \frac{|\theta^T s_b \theta|}{|\theta^T s_w \theta|} \quad (9)$$

Step 6 Apply the transformation matrix in (2) to get the low dimensional acoustic feature vector.

Experimental Setup

Speech Corpus

The ELDA Tamil Speech corpus was used in our experiment.80% of the data used for training and remaining 20% of the data used for testing. The parameter of the ELDA speech corpus is, sample rate-16kHz and 16 bit, wave format-mono, wav.

Feature Extraction

The feature used in this experiment is 12th order Mel Frequency Cepestral Coefficients (MFCC) and an energy Coefficient along with their first and second temporal derivatives [13]. Here frame length is 25ms and a frame shift is 10ms for extracting the feature. Each speech data is transformed into a sequence of feature vectors consisting of the Mel-Frequency Cepstral Coefficients (MFCCs).

Dimensionality Reduction

Compute Statistics

Statistics like mean vector and scatter matrices are computed from the feature files. The mean vectors are $\mu 1, \mu 2, \ldots\ldots$, μn for each class compute from equation .The scatter matrices are *sb*-scatter between-class and *sw*-scatter within-class compute from (6) and (8).

Transformation Matrix

With the mean vectors and the scatter matrices being computed, LDA finds the Eigenvectors $s - 1sw$, which together make up the optimal transformation.

Generate New Feature Files

Optimal transformation is known; by applying this optimal transformation to each file in training corpus low-dimensional acoustic feature can be obtained.

Acoustical Model and Decoding

Low-dimensional acoustic feature vectors are trained by using the HTK toolkit. The acoustical model is trained from the following steps.

Create Prototype Models

At first it is necessary to specify the design considerations i.e. the number of states, the transition matrix and the parameters for the observation pdfs [14][15]. This is done in form of an HMM prototype file.

Model Initialization and Re-estimation

In the next step it is necessary to find good estimates for the HMM model parameters. This is realized by using the tools HcompV, Hinit and Hrest The final result of this step are HMM definition files for all phonemes.

Decoding

Hvite is used for recognize the test data. The inputs to the Hvite are test data, HMM definition file, dictionary, network and list of phonemes in the HMM state. Hparse uses the language model and dictionary to produce the network file. As well as providing basic recognition, Hvite can perform forced alignments, lattice re-scoring.

Experimental Result

In our research two experiments can be carried out to show the effectiveness of linear dimensionality Reduction technique. These two experiments were done with ELDA Tamil speech corpus. First experiment uses high dimensional acoustic feature vector and second experiment was done with low dimensional acoustic feature vector. The Table1 shows the comparison of two experiments.

Table 1 Comparison between high dimensional acoustic vector and low dimensional acoustic vector.

Techniques	Speech Corpus	WER
High Dimensional Acoustic Feature Vector	ELDA corpus	15.4
Low Dimensional Acoustic Feature Vector	ELDA corpus	15.6

CONCLUSION

This paper presents a method of dimensionality reduction in Tamil language speech recognition system using Linear Discriminant Analysis. LDA was applied successfully to find an optimal linear combination of successive vectors of a feature stream for automatic speech recognition. LDA reduces the training time of the acoustical model also reduces the size of the acoustical model. This paper concludes Linear discriminant analysis (LDA) is a good methodology in finding an optimal linear feature subspace. LDA produces better result than the other dimensionality reduction techniques.

Acknowledgment

We would like to thank Evaluations and Language resource Distribution Agency for providing the ELRA: S0205 corpus.

References

1. Xuechuan Wang.l, Kuldip K.Paliwal, Feature extraction and dimensionality reduction algorithms and their applications in vowel recognition,,2003 Pattern Recognition Society, Elsevier Ltd.Pageno 2429 - 2439.
2. R. Haeb-Umbach, and H. Ney, Linear Discriminant Analysis for Improved Large Vacabulary Con- tinuous Speech Recognition,, in Proceedings of the IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, Vol. 1, pp. 13-16, 1992.
3. Arkadiusz Nag´orski , Lou Boves , Herman Steeneken, Optimal Selection Of Speech data for automatic speech recognition systems,, Inter speech-2002.
4. Arvind Agarwal, Jeff M. Phillips, Suresh, *Universal Multi-Dimensional Scaling*, ACM 978-1-4503- 0055-110/07- July 25-28, 2010.
5. S. T. Roweis and L. K. Saul, Nonlinear Dimensionality Reduction by Locally Linear Embedding, Science Vol 290, 22 December 2000, 2323-2326.
6. Haeb-Umbach,R., *Linear discriminant analysis for improved large vocabulary continuous speech recog-nition*,, IEEE International Conference on 23-26 Mar 1992.
7. Aleix M. Martinez, Aleix M. Mart'inez , Avinash C. Kak ,*PCA versus LDA*,,IEEE Transactions on Pattern Analysis and Machine Intelligence -2001.
8. Shiqing Zhang, Dimensionality reduction –based phoneme recognition,, IEEE conference Oct 2008. [9] Nagendra Kumar, Investigation of Silicon Auditory Models and Generalization of Linear Discriminant Analysis for Improved Speech Recognition,, Ph.D. thesis, Johns Hopkins. University, 1997.
9. Mark Hasegawa-Johnson, Feature Reduction with Linear Discriminant Analysis and its Performance on Phoneme Recognition,, May 8, 2004.
10. Jieping Ye, Ravi Janardan, Qi Li, Two-Dimensional Linear Discriminant Analysis, *Journal of the American Statistical Association.* 97(457):77-87, 2002.
11. Lindasalwa Muda, Mumtaj Begam and I. Elamvazuthi, Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques, *Journal of Computing*, Volume 2, Issue 3, March 2010.
12. Rabiner L.R, A tutorial on hidden Markov models and selected applications in speech recognition, IEEE-1989.
13. Steve Young, The HTK Book,, December 1995.

How to cite this article:

Sundarapandiyan S and Shanthi N (2017) 'Automatic Speech Recognition System of Tamil Language Using Linear Discriminant Analysis ', *International Journal of Current Advanced Research*, 06(09), pp. 6298-6301.
DOI: http://dx.doi.org/10.24327/ijcar.2017.6301.0913

****