A Novel Techniques for speech recognition using modified MFCC

Mohit Kumar Goel
Lovely Professional University, School of Electrical and Electronics Engineering, Jalandhar, India

Abstract

Recent studies have been focusing upon the performance of the speech recognition systems. The performance of robust speech-recognition system is measured in terms of accuracy of identification, computational cost and calculation speed. In speech recognition process, feature extraction is of utmost importance. Therefore, in this context, a feature extraction technique i.e. MFCCs has been discussed. Based on this, a technique has been proposed which combines conventional MFCC technique with Laplacian Eigenmaps. This proposed technique leads to dimensionality reduction of the feature matrix obtained by MFCC method, thus improving the performance system in terms of computational cost.

Keywords: Speech recognition, MFCC, Laplacian Eigenmaps, Feature extraction, dimension reduction.

I. INTRODUCTION

Speech recognition is a process of recognition of phonemes, words or sentences uttered by the person. In the past few years, lots of advancements have been made in the field of Speech Recognition Systems. The first and foremost step of Speech Recognition System is feature extraction. This is done as our speech signal contains undesired or redundant information. So, rather than processing the whole audio signal, certain features are extracted using different techniques. There are lots of techniques of feature extraction such as: Linear Prediction coding (LPC), Mel Frequency Cepstral Coefficients (MFCC), Perceptual Linear Prediction (PLP), Relative Spectral –PLP (RaSta-PLP), Principle Component Analysis (PCA), Kernel-based PCA (KPCA) and so on. Out of these, MFCC method has been used widely in many applications. This is because MFCC uses Mel scale which approximates the human auditory system in a better way as compared to all other techniques. Data obtained after feature extraction has high dimension which can reduce the performance of the speech recognition system. So this problem can be solved by using dimensional reduction techniques such as Principle Component Analysis, Laplacian Eigenmaps etc. PCA is linear dimensionality reduction technique in which the variance of the data is preserved. It is a non-linear dimensionality reduction algorithm which preserves the local structure of the dataset. In this algorithm, neighborhood graph is created or distance between the neighbors are calculated which helps it to retain the local information and makes it insensitive to the noise. Thus by using these techniques in combination with MFCC improves the performance of the speech recognition system in terms of reduction in noise and computational complexity [1].

In this study, feature extraction and dimensional reduction techniques have been discussed. In addition to this, a proposed method which combines MFCC with Laplacian Eigenmaps has been discussed. Moreover, the performance of the conventional MFCC method and proposed method has been compared in terms of computational cost.

II. Traditional MFCC method

MFCC uses mel scale which is based on the human ear or perception. It relates pitch of a tone to its actual frequency. In this, the frequency bands are positioned logarithmically on the mel scale, therefore, this approximates the human auditory system’s response more closely than the linearly spaced frequency bands of FFT. The basic block diagram of MFCC feature extraction technique is shown in Fig 1:
a. **Pre-emphasis**

The first step is to do Pre-emphasis. Pre-emphasis means applying High Pass filter to the speech signal to enhance the high frequency component of speech signal \( s(n) \). It is done using equation:

\[
\hat{s}(z) = 1 - a z^{-1}
\]

where \( a = 0.95 \) or \( 0.97 \)

b. **Framing**

After Pre-emphasis, the speech signal is divided into frames. Due to the randomness or non-stationary nature of the speech signal, it is difficult to recognize it which makes it harder for hardware implementation. Therefore, the speech signal is divided into small frames, and then each frame is spectrally analyzed.

c. **Windowing**

The commonly used window is hamming window. Therefore, hamming window is applied to each frame, thus decreasing the discontinuities in the beginning and end of the frame [base paper]. The expression for Hamming window is given as:

\[
w(n) = 0.54 - 0.46 \cos \left( \frac{2\pi n}{k} \right)
\]

Here \( k \) is the window length.

d. **Fast Fourier Transform (FFT)**

Now, the frequency spectrum of the speech signal is obtained by applying Fourier transform on the windowed signal. This is done because the human cochlea vibrates depending on the frequency of the incoming sounds. Therefore, on the basis of location in the cochlea that vibrates, different nerves send this information to the brain about the occurrence of certain frequencies. Hence, in the similar way, the spectrum identifies which frequencies are present in the frame.

e. **MEL frequency Warping**

Since, still the spectrum contains the irrelevant information; therefore, clumps of spectrum bins are taken and summed up to determine the amount of energy that exists in different frequency regions. Hence for this, MEL filter bank is created and the frequency spectrum is multiplied with this filter bank. The spacing of this filter bank is based on mel scale. The linear frequency scale is converted to mel-scale using equation:

\[
m = 2595 \times \log \left( 1 + \frac{f}{700} \right); \text{ where } f=\text{linear frequency.}
\]
f. Logarithm

After creating MEL-filter bank, logarithm of the mel-created energies is taken. This logarithmic operation is performed to compress higher frequencies as human ear is insensitive to the higher frequencies. By doing this, the features match more closely what humans actually hear.

III. Proposed method: Combination of MFCC and Laplacian Eigenmaps

After the application of conventional MFCC method and getting a feature matrix, a dimension reduction technique is applied. Since digital audios or speech signals are of high dimensionality [3], there exists redundancy in the features of the audio signals. Therefore, this redundancy is likely to be reduced so as to save memory and to reduce the transmission cost. Moreover, in the process of feature extraction, this leads to the compact representation of the data set with the reduction of noise.

Hence, in this work, Laplacian Eigenmaps is applied. It is a non-linear dimensionality reduction algorithm which preserves the local structure of the dataset. In this algorithm, neighborhood graph is created or distance between the neighbors is calculated which helps it to retain the local information and makes it insensitive to the noise. It is based on the Laplace-Beltrami operation to obtain the eigen functions.

The steps of Laplacian Eigenmaps algorithm are given as:

a. Neighborhood definition or construction of adjacency graph:

In this, adjacency matrix is obtained using either k-neighborhood or ε-neighborhood on data X. In order to construct the adjacency matrix, element (i,j) of the matrix is set to 1 if node i is connected to node j, and 0 otherwise.

b. Weighted graph creation:

The adjacency matrix obtained in step 1 can be considered as weight matrix, so it will not have parameter. Weight matrix can also be created using the formula as given below:

\[ W = \exp(-d^2/t) \]

where t is kept \( \infty \).

For the parameter \( t > 0 \), the above expression gives \( W = 1 \); otherwise it is set to 0.

c. Kernel Construction:

Let \( d_i = \sum w_{i,j} \) and the diagonal matrix be denoted as:

\[ D = \text{diag}(d_1, \ldots, d_n) \]

Therefore, the kernel is given as:

\[ L = D - W \]

d. Eigen Decomposition of the kernel:

The eigen decomposition is done using the equation as given below:

\[ Lf = \lambda Df \]

Here \( \lambda \) denotes the eigen values for \( i = 0 \) to \( n-1 \) and \( F = \{ f^1, \ldots, f^n \} \) denotes the harmonic eigen vector matrix. Therefore, in Laplacian eigen maps algorithm, the eigen vector corresponding to 0th eigen value has been neglected and rest all are considered into account to get reduced matrix. Hence, the required reduced matrix is obtained as:

\[ Y = F' \]
The block diagram of this proposed method is shown as below:

![Block Diagram](image)

After applying Laplacian Eigenmaps to the MFCC feature matrix, its dimension is reduced.

### IV. Experimental Set up

The speech recognition system consists of two phases: One is training phase and other is testing phase.

**a. Training Phase:**

In training phase, firstly the speech samples or utterances are recorded to prepare the database. The database used in this work consists of 300 utterances of 30 different words from 10 different speakers. After preparing database, following steps are followed.

#### i. Feature Extraction using MFCC method

The speech signal is non-stationary in nature. Because of this, speech signal is divided into frames and then each frame is analyzed separately. The basic steps of MFCC algorithm are explained in section II of this paper.

#### ii. Dimensionality Reduction method

After getting the feature matrix of given speech sample, the dimensionality reduction technique i.e. Laplacian Eigenmaps technique is applied. Hence, after applying Laplacian Eigenmap method on each utterance of database, the size of matrix of each utterance is $(11 \times 12)$ or $(1 \times 132)$.

#### iii. Clustering of the data

After getting the reduced matrices for each utterance of database, the matrices are saved as excel file (.xls file). After this, these spreadsheets are imported into a single matrix say S. Therefore, the size of matrix S is $300 \times 132$. Then, on this data matrix S, K-mean algorithm is applied in order to cluster the data into K-clusters. Therefore, on applying K-mean algorithm,
the centroid location for each utterance matrix is obtained with the cluster indices. Hence with this, training phase is complete.

b. Testing Phase:

During testing phase, some utterances from the complete database are used for testing. For testing, all the steps are followed as in case of training till getting Laplacian matrix for respective testing utterance. Then the comparison is done between the testing sample and the centroids obtained after K-mean clustering. For this, the Euclidean distance is calculated. As a result of this, the cluster with minimum distance is obtained which contains that speech sample corresponding to the testing utterance. Therefore, again the comparison is made between the testing utterance and the samples contained in that cluster. Therefore, the one which is having minimum distance from all is considered as the result. Hence by using K-mean algorithm, the number of comparisons is reduced since instead of making comparison with the complete database, the comparison between the utterances contained only in the cluster having minimum distance and testing utterance is done.

V. Experimental Results

By applying the dimensionality reduction technique after extracting feature matrix, the dimensions are much reduced. For example, the MFCC matrix for the word “beat” has size \((22 \times 12)\). After applying Laplacian Eigenmaps, the dimension is reduced to \((11 \times 12)\). Therefore, this is applicable for all the 300 utterances. Hence the size of the resultant matrix containing all the 300 matrices as sub-matrices is reduced. Due to this, the number of calculations is reduced and thus, computational cost is reduced. For example: there are 300 words in database, therefore, the number of calculations are reduced by 56%. Hence, Laplacian Eigenmaps leads to reduction in computational cost.

<table>
<thead>
<tr>
<th>Technique</th>
<th>Computational Cost per word</th>
</tr>
</thead>
<tbody>
<tr>
<td>MFCC (for 300 words in database)</td>
<td>300</td>
</tr>
<tr>
<td>Laplacian Eigenmaps (for 300 words in database)</td>
<td>132</td>
</tr>
</tbody>
</table>

In addition to this, Laplacian Eigenmaps method also provides improved recognition accuracy as compared to Principle Component Analysis method which is also a dimensionality reduction technique. Both methods are applied to MFCC feature matrices and results are compared.

![Fig 4: Recognition accuracies of PCA and Laplacian Eigenmaps](image-url)
VI. CONCLUSION

Dimensional reduction method, i.e. Laplacian Eigenmaps, is used to reduce the size of the original set of features by retaining as much information as possible, thus solving the problems caused by high dimensionalities. Therefore, this method is applied after the feature extraction using MFCC. Due to the reduction in dimension of the feature matrix, this leads to less use of memory space, faster calculation while testing, therefore, resulting in less computational cost as compared to MFCC method. Hence it is observed that Laplacian Eigenmaps leads to reduction in computational cost and also the recognition accuracy is improved as compared to PCA.

REFERENCES

