International Journal Of

Recent Scientific Research

ISSN: 0976-3031

Volume: 7(3) March -2016

LANGUAGE TRANSCRIPTION AND TRANSLATION BY USING DEEP NEURAL NETWORK (DNN) BASED ON WAVELET TRANSFORM

Sona S., Swetha S., Vaishnavi G and Srinath R

THE OFFICIAL PUBLICATION OF
INTERNATIONAL JOURNAL OF RECENT SCIENTIFIC RESEARCH (IJRSR)
http://www.recentscientific.com/ recentscientific@gmail.com

ISSN: 0976-3031

Available Online at http://www.recentscientific.com

International Journal of Recent Scientific Research
Vol. 7, Issue, 3, pp. 9347- 9351, March, 2016

International Journal of Recent Scientific Research

RESEARCH ARTICLE

LANGUAGE TRANSCRIPTION AND TRANSLATION BY USING DEEP NEURAL NETWORK (DNN) BASED ON WAVELET TRANSFORM

Sona S., Swetha S., Vaishnavi G and Srinath R*

ARTICLE INFO

Article History:

Received 15th December, 2015
Received in revised form 21st January, 2016
Accepted 06th February, 2016
Published online 28th March, 2016

Keywords:

Deep neural networks(DNNs),Gaussian mixture model(GMM), Senones, Context –dependent deep neural network(CD-DNN),HAAR transform, Feed forward network

ABSTRACT

Deep neural networks(DNN) and Gaussian mixture model(GMM) has recently achieved significant performance gain and better efficiency in spoken language recognition(SLR) which identifies the language being spoken. But in our system we propose language translation and transcription of input signal (human voice) which produces output in text form of a standard language depending on the user's application area. Senone posteriors used in context- dependent deep neural network (CD-DNN) which recognises the language spoken by people irrespective of pronunciation. Wavelet (HAAR) transform is applied on voice signal to obtain features of voice. Feed forward network issued to pass the input signal to hidden layer of NN network and obtain the primitive function of node. Computational speed of 1-3 sec is achieved, Significant improvement in gain and efficiency of (65-73%) is obtained compared to Gaussian mixture model (GMM).

INTRODUCTION

The introduction of Deep Neural network (DNN) had a substantial impact on recognizing multilingual speech [7]. In many language –related problems a language model is the statistical model that acquires individual words from which meaningful sentences can be constructed.

Using factorization of Baye's division rule, the language model [LM] estimates the probability of occurrence for a given word sequence in neural network and count LM usually do not exceed a content size of 3-4 words and increasing

Computational complexity in neural network LM while decoding [3]. The existing LM has two main drawbacks in constructing larger language model:i) It is not convenient to obtain larger language words using a single DNN ii)Most of the previous studies deals with the monolingual information from the direct words only[8].

Our recent study investigates a feature based technique to improve language transcription and translation (LTT) performance among various kinds of noise. In contrast to adaption model feature based LTT technique keep the LTT backend unchanged. During transcription and translation feature enhancement made to control noise [2]. In this paper we have overcome the drawback of language model to predict the exact senone posteriors irrespective of speaker pronunciation .Here we are taking four languages for extracting the features of the input voice signal.

Usually mean and covariance[1] is used for extracting the features from input voice signal .The input voice signal which has more front-end noise. To get rid of front-end noise, Ratio of time -frequency masking plays an important role in DNN but it increases more complexity [2]. To reduce the complexity and error rate, correlation is obtained in feature extraction to eliminate the front-end noise.

The DNN approach plays a crucial role in speech recognition. The fusion system of DNN includes two different DNN's where the two DNN's are stacked: the bottle neck features from the first DNN over a context of 10 frames are used as input to the second DNN's with similar construction as the previous one. In this the author [1] gained 4 to 10% in 3 to 10 sec condition using single DNN approach.

By senone posteriors exact corpora is obtained despite the varying pronunciation. In order to achieve more gain, 4frames are stacked continuously to obtain the exact transcripted sentences.

*Corresponding author: **Srinath R**

The current work presents the language translation and transcription process where data from four languages namely English, Hindi, Tamil and Telugu are obtained and stacked in 4 frames of DNN layer. Feed forward network is used to pass the input signal to hidden layer of Neural network (NN) and obtain the primitive function of node. Back propagation model is used to adjust the weights and biases of networks to minimize the sum squared error of the network which gives fast and accurate classification. Back propagation algorithm finally classifies the pattern of characters.

Background and System Description

The following section describes how input signals are preprocessed to obtain features of applied signal. In contrast to that the features are extracted from database signal being stored in stacked DNN layer which are subjected to Deep learning. The input voice signal is classified and finally language translation and transcription is achieved.

Wavelet Transform

Discrete wavelet transform (DWT) is a popular method used for transformed feature extraction of a pre-processed signal. It allows decomposition of a signal with wavelet in the form of a low pass and high pass filter signal. In DWT the wavelets are discretely sampled.

Compared to other transforms DWT has an advantage that it is localised in both time and frequency. Signal de-noising is better in DWT compared to other transforms which removes noise and signal spectra overlap.

Haar wavelet transform is the simplest possible wavelet transform. Haar wavelet is a sequence of rescaled "Square shaped" function which together form a wavelet family that allows a target function over an interval to be represented in terms of an orthonormal basis. The technical disadvantage for analysis of signals with sudden transition.

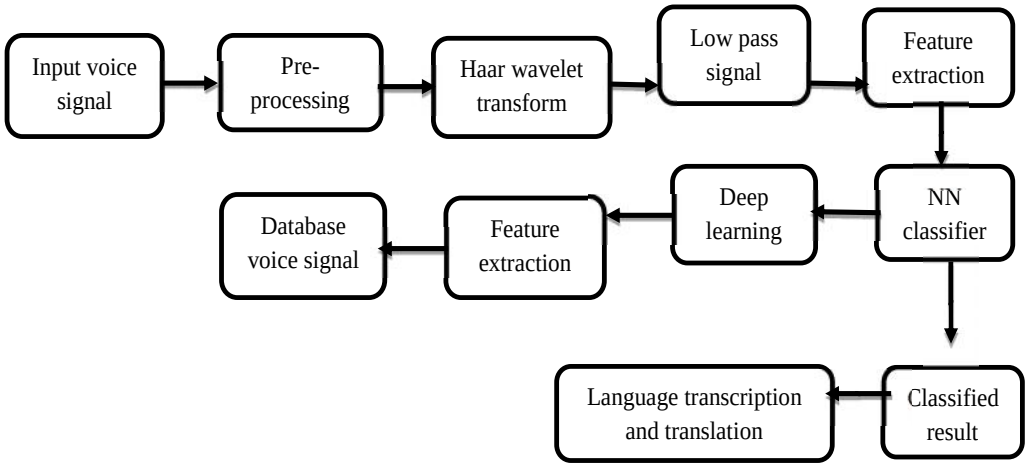

Fig.1 Block diagram for language transcription and translation using NN classifier

Pre –Processing

The human voice signal is applied as input which basically contains noise and must be subjected to pre-processing. The noise present in the signal includes Gaussian noise and environmental noise. The input voice signal being fed to the model is is from any of the four languages namely English, Hindi, Tamil and Telugu.

The pre-processing stage in speech recognition system is used to increase the efficiency of subsequent feature extraction of signal in order to improve the performance. Pre-processing contains three stages. In the first stage sampling is performed to satisfy the Nyquist sampling rate.

In the second stage of pre-processing median filtering technique is applied to eliminate the variations and disproportions in the spread of the signal. The final stage of pre-processing contains de-noising to obtain a noise free speech signal.

For an input represented by a list of 2^n numbers, the input values are paired up storing difference and passing the sum. This process is repeated recursively, pairing sum to provide next scale, which leads to 2^{n-1} difference and sum. A high frequency sub band contains edge information of input signal and low sub band contains clear information about signal. The low pass filtered information is obtained after wavelet transformation and passed to consecutive stages.

Feature Extraction

To obtain the exact results of senone posteriors from the applied voice signal the following features are computed which are explained as follows.

Mean

The mean indicatedby μ used to compute the average value of a signal. It is followed by adding all of samples together and divide by N. The signal is contained in x_o through x_{n-1}, i is an index that runs through these values.

$$\mu = \frac{1}{N}\sum_{i=0}^{N-1} x_i$$

Entropy

Entropy returns a scalar value representing the entropy of low pass signal. Entropy is a statistical measure of randomness that can be used to characterize the low pass signal. It quantifies the capacity of information source.

The calculation of entropy of speech carry various form of information as phenomes, intonation signals, accent, speaker voice and speaker stylistics.

$$E(X) = -\sum_{i=1}^{M} P_x(x_i) \log P_x(x_i)$$

Variance

Variance gives the measure of how the data signal distributes itself about the computed mean or expected value. Variance is always non-negative .Small variance indicates data points tends to be close to mean, and those which has a high variance value indicates that data points are spread out around mean from each other.

$$\sigma^2 = \frac{1}{N-1} \left[\sum_{i=0}^{N-1} x_i^2 - \frac{1}{N} \left(\sum_{i=0}^{N-1} x_i \right)^2 \right]$$

The signal is expressed in terms of three accumulated parameters: N represents the total number of samples, sum represents the sum of these samples, and sum of squares of these samples.

Standard Deviation

Standard deviation (σ) is a measure of the spread of scores within a set of data. It is generally the square root of variance. It is a measure of how far the signal fluctuates from its mean.

$$\sigma = \sqrt{\left(\frac{1}{N-1} \left[\sum_{i=0}^{N-1} x_i^2 - \frac{1}{N} \left(\sum_{i=0}^{N-1} x_i \right)^2 \right] \right)}$$

Correlation

A Distribution involving two variables such that change in one variable affects the change in other variable, the variables are said to be correlated. It refers to measuring the closeness of relationship between the variables.

$$r = n(\sum xy) - (\sum x)(\sum y) / \sqrt{[n\sum x^2 - (\sum x^2)][n\sum y^2 - (\sum y^2)]}$$

Deep Learning

In Database about 50 sentences from four different languages are chosen and features (Mean, Entropy, Variance, Standard Deviation and Correlation) are computed. Thus 1000 features are obtained which are stored in subsequent DNN layers where they are stacked upon each other. A total of 200 DNN layers are used for storing the database signal. The sentences which are stored in NN layer are subjected to deep learning which requires 15 hours of training.

Feed forward network is used where the input signal is fed into the network. The primitive function and their derivatives are evaluated at each node. The derivatives are stored .Back propagation learning rule can be used to adjust the weights and biases of networks.

The constant 1 is fed into the output unit and the network is run backwards. Incoming information to a node is added and the result is multiplied by the value stored in the left part of the unit. The result is transmitted to the left of the unit. The result collected at the input unit is the derivative of the network function with respect to input signal.

RESULTS

The waveforms are obtained for the sample input voice signal applied and the processing of signal in consecutive stages in block diagaram is shown as follows:

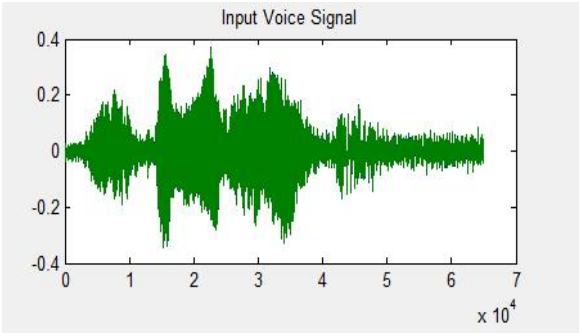

Fig 2 Waveform of input voice signal

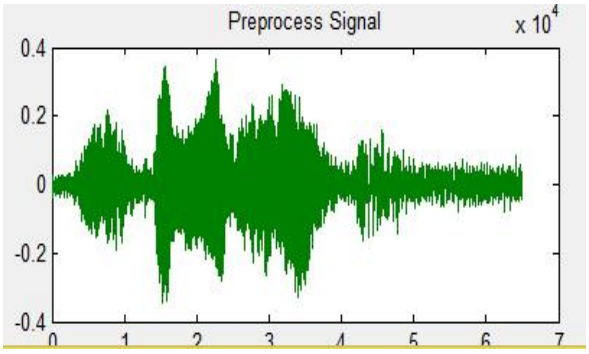

Fig 3 Waveform of a pre-processed signal

In the above fig .2 the typical waveform of an input human voice signal is shown. The input speech signal is contaminated with noise and the noise must be segregated and hence subjected to pre-processing. The Pre-Processed signal removes the noise present in the input signal and is shown in Fig.3.

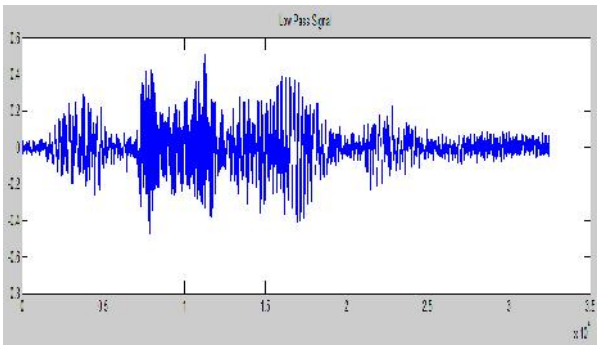

Fig. 4 Wavelet transform of filtered low pass signal

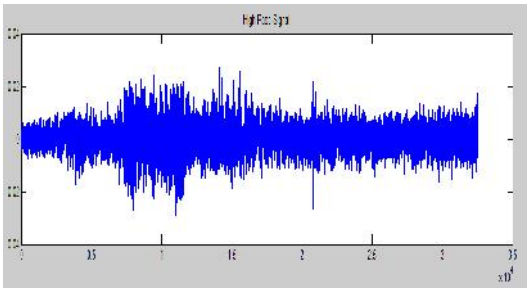

Fig.5 Wavelet transform of a filtered high pass signal.

The pre-processed signal is subjected to Wavelet transform which divides the signal into filtered low pass and high pass signal as shown in Fig .4 and Fig.5 respectively.

A high pass filtered signal usually contains the edge information about the signal whereas a low pass signal contains the clear information about the signal. Hence a low pass signal is taken and features are extracted for a low pass signal.

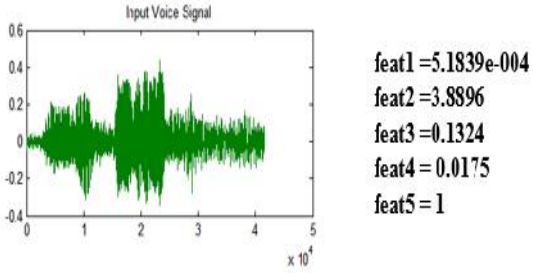

Fig 6 Input signal and feature extraction for Tamil language.

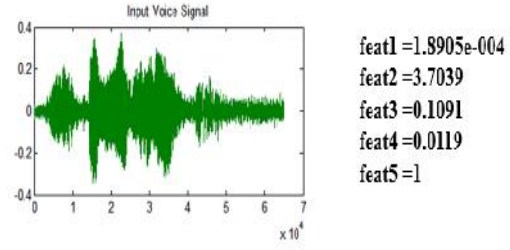

Fig 7 Input signal and feature extraction for telugu language

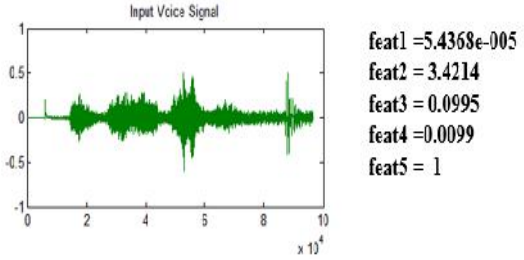

Fig 8 Input signal and feature extraction for Hindi language

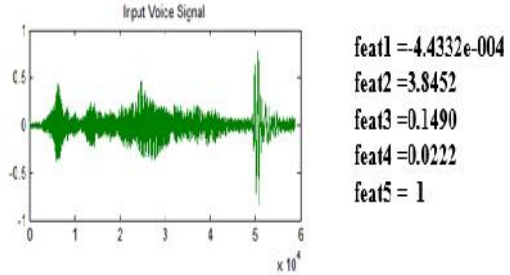

Fig 9 Input signal and feature extraction for English language

CONCLUSION

The applied input signal features are compared with the database signal features, where the collected voice signals are stored in subsequent layers of neural network. The signals are subjected to deep learning. The features of input signal and Database signal are compared by using back propagation algorithm which gives faster classification and better accuracy. The training of data samples require large amount of time but computational speed and efficiency is fast of the order of 1-3 sec. Since wavelet transform is used, a noise free signal is obtained. Computing features for each speech signal gives clear details about the signal.

By using Deep Neural Networks a large amount of database can be stored .Accuracy is more .Thus faster classification of language is obtained and the outcome includes Language transcripted and translated voice along with text display which has proven better performance compared to previous models.

References

1. Study of senone -based Deep neural Network Approaches for Spoken Language Recognition Luciana Ferrer, Yun Lei, Mitchell McLaren, and Nicholas Scheffer, Vol.24, No.1, Jan 2016.
2. Improving Robustness of Deep Neural Network Acoustic Models via Speech Separation and Joint Adaptive Training Arun Narayanan, Student Member, IEEE, and DeLiang Wang, Fellow, IEEE Vol.23, No.1, Jan 2015.
3. From Feedforward to Recurrent LSTM Neural Networks for Language Modelling Martin Sundermeyer, Hermann Ney, Fellow, IEEE, and Ralf Schluter, Member, IEEE Vol.23,No.3, Mar 2015.
4. State-Clustering Based Multiple Deep Neural Networks Modeling Approach for Speech Recognition Pan Zhou, Hui Jiang, Senior member, IEEE, Li-Rong Dai, Yu Hu, and Qing-Feng Liu Vol.23,No.4,Apr 2015.
5. Speaker and Expression factorization for Audiobook Data: expressiveness and transplantation Langzhou Chen, Member, IEEE, Nobert Braunschweiler, and Mark J,F. Gales, Fellow, IEEE Vol.23,No.4,Apr 2015.
6. Deep Learning for Acoustic modelling in parametric Speech Generation Zhen-Hua Ling, Shi-Yin Kang, Heiga Zen, Andrew Senior, Mike Schuster, Xiao- Jun Qian, Helen Meng, and Li Deng IEEE signal Processing Magazine May 2015.
7. A Real-Time End-to-End Multilingual Speech Recognition Architecture Javier Gonzalez-Dominguez, Member, IEEE, David Eustis, Ignacio Lopez-Moreno Member, IEEE, Andrew Senior, Senior Member, IEEE, Francoise Beaufays, Senior Member, IEEE, and Pedro J. Moreno, Senior Member, IEEE Vol.9, No.4, Jun 2015.
8. Bilingual Continuous-Space Language Model Growing for Statistical Machine Translation Rui Wang, Hai Zhao, Bao-Liang Lu, Senior Member, IEEE, Masao Utiyama and EiichiroSumita Vol.23, No.7, July 2015.

9. Multitask Learning of Deep Neural Networks for Low-Resource Speech Recognition Dongpeng Chen and Brian Kan-Wing Mak Vol.23, No.7, Jul2015.
10. Use of Micro-Modulation Features in Large Vocabulary Continuos Speech Recognition Tasks Dimitrios Dimitriadis, Senior Member, IEEE, and Enrico Bocchieri, Fellow, IEEE Vol.23, No.8, Aug 2015.
11. Data Argumentation for Deep Neural Network acoustic Modeling Xiaodong cui, Senior Member, IEEE, VaibhavaGorel, Senior Member, IEEE, and Brian Kingsbury, senior Member, IEEE Vol.20, No.9, Sep 2015.
12. Deep Neural Networks for Single-Channel Multi-Talker Speech Recognition Chao-Weng, Student Member, IEEE, Dong Yu, Senior Member, IEEE, Michael L.Seltzer, Senior Member, IEEE, and Jasha Droppo, Senior Member, IEEE Vol.23, No.10, Oct 2015.
13. Deep Neural Network Approaches to Speaker Recognition and Language Recognition Fred Richardson, Senior Member, IEEE, Douglas Reynolds, Fellow, IEEE, and NajimDehak, Member, IEEE Vol.22, No.10, Oct 2015.
14. Morphologically Filtered Power-Normalized Cochleograms as Robust, Biologically Inspired Features for ASR Fernando de-la-calle-Silos, Student Member, IEEE, Francisco J. Valverde-Albacete, Member, IEEE, Ascension Gallardo-Antolin, and Carmen Pelaez-Moreno, Member, IEEE Vol.23, No.11, Nov 2015.

****

How to cite this article:

Sona S., Swetha S., Vaishnavi G and Srinath R.2016, Language Transcription And Translation By Using Deep Neural Network (Dnn) Based On Wavelet Transform. *Int J Recent Sci Res.* 7(3), pp. 9347- 9351.

TSSN 0976-3031