International Journal of Emerging Research in Management &Technology ISSN: 2278-9359 (Volume-6, Issue-9)

Research Article | **September 2017**

New Feature Vectors for Language Identification Using Deep Neural Networks

Dr. A. Nagesh

Professor, Department of Computer Science & Engineering, Mahatma Gandhi Institute of Technology (MGIT) Jawaharlal Nehuru Technological University (JNTUH) Hyderabad, Telangana, India

Abstract:

The impressive performance of neural networks (NNs) for automatic speech recognition has motivated us to use for language identification (LID). In this paper, a new features based language identification system using neural network is presented. The new feature vectors are extracted based on the principle the frequency of occurrence phonemes is different among the languages. In this new form of feature vectors, the feature vectors are represented as a probability vector instead of scalar value. Because of this these new form of feature vectors, the DNN classifier classify the languages under consideration accurately.

Keywords: Language Identification (LID), Deep Neural Networks (DNN), Mel Frequency Cepstral Coefficients (MFCC).

I. INTRODUCTION

Automatic language identification (LID) is the process by which the language of a spoken utterance is identified. There is a variety of information that human and machines can use to distinguish one language from another. Generally, different speech information for LID task can be divided into the spoken level and word level.

Recent findings in the field of language identification have shown that significant accuracy improvements over classical GMM method can be achieved through the use of neural network (NN). Motivated by the recent success of using neural networks in acoustic modeling for speech recognition, the NNs is used to identifying the language in a given utterance from its short-term acoustic features.

The Neural networks can be broadly classified into three categories, namely, feed forward neural networks (FFNN), feedback neural networks (FBNN) and the combination of both feed forward and feedback neural networks. In this study, we propose the use of neural networks (NNs) as a method to perform LID at the acoustic level. A Neural Networks are composed of simple elements operating in parallel. These elements are inspired by biological nervous systems. As in nature, the network function is determined largely by the connections between elements. We can train a neural network to perform a particular function by adjusting the values of the connections (weights) between elements. Commonly neural networks are adjusted, or trained, so that a particular input leads to a specific target output. The network has varying neurons input n, which receive input of new feature vectors. Number of hidden layer varies from 1 to 4 layers and number of neurons in each hidden layer varies from 10 to 50 neurons.

The first feasibility study of language identification task is done by Muthuswamy [1]. The language identification using phonotactic and prosodic features is described by Yagnanarayana and mary Leela [2]. The LID task using deep neural network is explored by Lopez-Moreno [3].

Remaining paper is organized as follows: Section 2 describes the new feature extraction method. Baseline method using neural network is presented in section 3 and Conclusion in section 4.

II. FEATURE EXTRACTION

Performance of the LID systems mainly depends on the acoustic information used to represent the language discriminative information and the modeling technique used to develop the LID systems. The language discriminating acoustic information is represented as a new feature vectors. The language discriminating information is effectively captured using Gaussians.

In this work a new form of feature vector representation is described. The feature vectors are represented in the form of Gaussians. Using Gaussian Mixture model (GMM) as a front end the feature vectors are extracted from the speech signal. For any system the basic requirement is to obtain the feature vectors from the speech signal. In the literature is found that some attempts are made to explore the new way of representing the feature vector based on the GMM feature extraction. The feature vectors are represented in probability vectors form.

Instead representing GMMs as scalar value, it is represented as a probability vector. So the system performance is improved. For LID task, the new feature vectors are obtained from the speech signal estimating using probability density function based on Gaussian mixture model. The underlying language specific discrimination information is represented as a Gaussians [4][5][6].

In this section a new form of feature vector extraction is described for LID task. Here the feature vectors are represented in the form of Gaussians. Using Gaussian Mixture Model as a front end the feature vectors are extracted from the speech signal. The first step in the language identification task is MFCC feature extraction from the speech signal. In the literature it is found that some attempts are made to explore the new way of representing the frequency of occurrence of phonemes is different between among the languages based on principle, the feature vectors are extracted using GMM feature extraction. The feature vectors are represented in probability vectors form, instead representing GMMs as scalar value it is represented as a probability vector. So the LID system performance is improved. For language identification task, the new feature vectors are obtained from the speech signal estimating using probability density function based on Gaussian mixture model. The underlying language specific discrimination information is represented as a Gaussians. Beginning from the training data of language L_i , a MFCC feature vectors are extracted with a frame size of 25ms and frame shift of 10ms. These feature vectors are grouped into clusters with „R" Gaussian mixtures as shown in Fig.1.

Fig.1: R Gaussians for language L_i.

2.1 Computation of New Feature Vectors

Once „R" Gaussians, R clusters are formed using MFCC based feature vectors. Each cluster represented as one Gaussian. The feature vector $X=(X_1,X_2,\ldots X_{12})$ is passed through a Gaussian G_1 by calculating probability P_1 using probability density function (PDF) of Gaussian G_1. This P_1 is first coefficient in the new feature vector. In the same way feature vector X is passed through

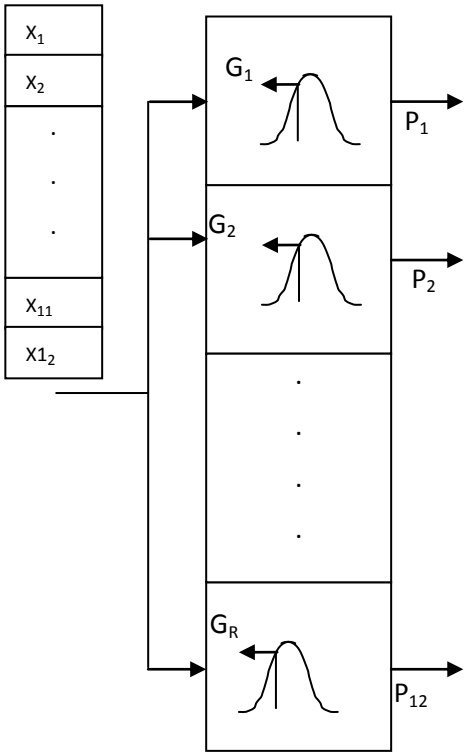

Fig.2: Parameter estimation for new feature vector P.

R Gaussians by creating R feature vector coefficients namely $P_1,P_2,\ldots,P_R$, as shown in Fig. 2. These R coefficients create a new feature vector of dimension R. As explained above, all the feature vectors are passed through R Gaussians (G_1, $G_2,\ldots$, G_R) generating new R dimensional feature vectors. In other words the 12 dimensional MFCC feature vectors of size N are transformed to R dimensional feature vectors of size N. The 12 dimensional MFCC feature vector is represented as a new „R" dimensional feature vector[7][8][9].

In the new feature vector, each Gaussian probability density represents one coefficient. Experiments are carried out to find the dimension of new feature vector for good language identification performance. This is done by varying the number of Gaussians (coefficients) from 10 to 15, i.e number of coefficients in the new feature vector. The newly derived feature vectors are given to the neural network based classifier for language identification [10].

III. BASELINE SYSTEM FOR LANGUAGE IDENTIFICATION USING ANN

The ANN is the most efficient technique for spoken language identification. The new feature extraction method is used for extracting features from the speech signal and ANN is used as the identification method. The first subpart of the neural network is the learning and the second is the identification. For learning or training we applied the back propagation learning algorithm. We adjusted the weight and threshold in learning phase, and saved into the database. In the identification phase, we used the database from learning algorithm to match the unknown speech signals.

3.1 Experimental Setup

These experimental results shown in Table 1 is using ANN with three hidden layers, 50 neurons and linear transfer function for 1st hidden layer, 40 neurons and tan-sigmoid transfer function for 2nd hidden layer, 30 neurons and tan-sigmoid transfer function for 3rd hidden layer and tan-sigmoid transfer function for output layer.

3.2 Performance Analysis

The experiments are carried using new feature vectors of size. The new feature vectors are trained using ANN with a the order of 10,11,12, 13,14 and 15 of feature vector size. Testing is carried out using speech duration such as 1sec, 3sec and 5 sec. The identification performance is shown in the table 1.

Table.1 The performance new feature vector of varying feature vector size

New feature vector order	Training Error %	Testing Error %
10	20	20
11	15	15
12	10	11
13	9	9
14	8	8
15	6	6

IV. CONCLUSION

In this paper the impressive performance of neural networks in automatic speech recognition is applied to language identification using new form of feature vectors. The new form of feature vectors captured language specific information effectively. The performance ANN based LID system is superior to GMM based LID system.

REFERENCE

[1] R.A. Cole, J.W.T. Inouye, Y.K. Muthusamy, and M. Gopalakrishnan, Language identification with neural networks: a feasibility study, in Communications, Computers and Signal Processing, 1989. Conference Proceeding.,

[2] M. Leena, K. Srinivasa Rao, and B. Yegnanarayana, Neural network classifiers for language identification using phonotactic and prosodic features, in Intelligent Sensing and Information Processing, 2005.

[3] D. Yu and L. Deng, "Deep Learning and its Applications to Signal and Information Processing, *IEEE*, vol. 28, no. 1, pp. 145–154, 2011.

[4] N. Jaitly, P. Nguyen, A. Senior, and V. Vanhoucke, Application of Pretrained Deep Neural Networks to Large Vocabulary speech recognition, in Proceedings of Interspeech 2012.

[5] Richardson, F., Reynolds, D., Dehak, N., 2015. A Unified Deep Neural Network for Speaker and Language Recognition

[6] McCree, A., Multiclass discriminative training of i-vector language recognition. In: IEEE Odyssey: The Speaker and Language Recognition Workshop. Joensu, Finland, 2014.

[7] G. Hinton, L. Deng, D. Yu, G.E. Dahl, A. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T.N. Sainath, and B. Kingsbury, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," Signal Processing Magazine, IEEE, vol. 29, no. 6, pp. 82–97, 2012.

[8] Y. Lei, L. Ferrer, A. Lawson, M. McLaren, and N. Scheffer, "Application of convolutional neural networks to language identification in noisy conditions," in Proc. Odyssey-14, Joensuu, Finland, June 2014.
[9] I. Lopez-Moreno, J. Gonzalez-Dominguez, O. Plchot, D. MartınezGonzalez, J. Gonzalez-Rodriguez, and PJ Moreno, "Automatic language identification using deep neural networks," in Proc. ICASSP, Florence, Italy, May 2014.
[10] Desai S, Black A W, Yegnanarayana B, Prahlad K 2010 Spectral mapping using artificial neural networks for voice conversion. IEEE Trans. Audio Speech Lang. Process. 18(5): 954–964.

