International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-6, April 2020

Language Identification based on Support Vector Machine using GMM Super vectors

A.Nagesh

Abstract: ***This paper proposes a novel approach that combines the power of generative Gaussian mixture models (GMM) and discriminative support vector machines (SVM). The main objective this paper is to incorporating the GMM super vectors based on SVM classifier for language identification (LID) task. The GMM based LID system to capture all the variations present in phonotactic constraints imposed by the language requires large amount of training data. The Gaussian mixture model (GMM)-universal background model (UBM) modeling require less amount of training data. In GMM-UBM LID system, a language model is created by maximum a posterior (MAP) adaptation of the means of the universal background model (UBM). Here the GMM super vectors are created by concatenating the means of the adapted mixture components from UBM. Then these super vectors are applied to a SVM for classification purpose. In this paper, the performance of GMM-UBM LID system based on SVM is compared with the conventional GMM LID system. Form the performance analysis it is found that GMM-UBM LID system based on SVM is performed well when compared to GMM based LID system.***

Keywords : Language Identification, Gaussian Mixture Model, Support Vector Machine, Universal Background Model.

I. INTRODUCTION

The LID is a task of identifying the language from short utterance of the speech. The conventional LID system is based on the GMM. The recent development in GMM language identification task is the combining the GMM-UBM MAP adapted mean super vectors with SVM to create a hybrid GMM-UBM LID system[1]. The common techniques used for language identification task includes acoustic and phonotactic features for modeling. To captures all the variations effectively present in phonotctic constraints imposed by the language to model using GMM approach require large number of mixture components. To build a GMM based LID system with large number of mixture components require large amount of training data. It is tedious and time consuming task to get the large amount of training data. This requirement can overcome by using GMM-UBM super vectors for creating the language specific GMM models. This is because, the GMM-UBM modeling approach require less amount of training data. So we are employing a GMM-UBM based modeling technique, which creates the MAP adapted GMM mean super vectors. The state of art classification approach is support vector machine (SVM) which is used widely for LID and speaker identification task [2].

Revised Manuscript Received on April 30, 2020.
***** Correspondence Author
Dr A.Nagesh*, professor in CSE at MGIT, Hyderabad,India.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC-BY-NC-NDlicense (http://creativecommons.org/licenses/by-nc-nd/4.0/)

Support vector machine approach is proven more effective for language identification classification task. The SVM maps the input space to high dimensional space in nonlinear way. Then classification technique is applied to this potentially high dimensional space[3][4]. Here for the language identification task, GMM-UBM super vectors are combined with the SVM modeling to create a new hybrid LID system. The main idea in this new approach is that first set of all language is trained by UBM model and creating the language specific GMM by utilizing the current language utterance through maximum a posteriori (MAP) adaptation. In this adaptation only means of the mixture components is adapted and it is represented as GMM-UBM super vectors. Then these concatenated super vectors are given to SVM for the classification purpose (language identification purpose). In this paper, we propose a SVM based language identification system using GMM-UBM super vectors and evaluate the performance evaluation this LID system when compared to conventional GMM LID system[5].
The paper is organized as follows. First two explains the some aspects of GMM-UBM Super vector LID system based on SVM . The three section deals conventional GMM LID system. Section four contain the description GMM-UBM LID super vector SVM LID system. Followed by section discuss creation of GMM LID system, GMM-UBM LID system based on SVM and performance evaluation of above two LID systems.

II. GMM-UBM SUPER VECTORS LID SYSTEM BASED ON SVM

In the conventional GMM LID system, the identification performance of the system suffers due the variability of channel, speaker and environment present in the speech data. But the GMM-UBM based LID system is vulnerable to undesired variability due to non-language effect, such as channel, environment and speaker present in the speech data. This is because of the language specific GMM models are created using adapted mean UBM-GMM super vectors. The GMM-UBM super vectors are proven that it is very effective for language identification task.

A. Gaussian Mixture Model

GMM is a parametric estimation approach is used to estimate probability density function (pdf) from a set of feature vectors $X = \{x_1, x_2, , x_n\}$. The pdf is represented as a linear combination of M Gaussian components. The probability density of the d-dimensional feature vector is calculated from the function

$$p(x) = \sum_{i=1}^{M} p_i b_i(x)$$

Where p_i represents the mixture weight which satisfy the conditions such that

$\sum_{i=1}^{M} p_i = 1$ and $0 \leq w_i \leq 1$.

$p_i(x)$ indicate the Gaussian density function and is given by

$$b_i(x) = \frac{1}{(2\pi)^{D/2}|\sum_i|^{1/2}} e^{\left\{-\frac{1}{2}(x-\mu)^T \Sigma_i^{-1}(x-\mu)\right\}}$$

in the above equation μ_i is the N-dimensional mean vector and Σ_i is NXN dimensional covariance matrix. Here we consider only diagonal covariance. GMM model is resented as a $\lambda = \{p_i, \overline{\mu}_i, \Sigma_i\}$. The main objective is to estimate the GMM model parameters for feature vectors. The most preferred estimation approach is maximum likelihood (ML) estimation. The main criteria of ML estimation is to estimate the GMM parameters which maximize the given feature vectors. The ML GMM parameters are evaluated using Expected-maximum algorithm. The main idea of expected-maximization process is to start with a initial model λ and estimate a new model which satisfy the $(X/\lambda) < p(X/\hat{\lambda})$. the next iteration the new model $\hat{\lambda}$ model becomes the initial model and the process is repeated until the a convergence is reached.

B. GMM-UBM Super Vectors

However, it is not efficient to build one GMM for each language when there is no enough training data is available. Instead of creating one GMM per language, a language-independent universal background model is first created. For this with the set of all language data, it is modeled by using GMM to create the universal background model (UBM). From this model using language specific speech data the GMM model is adapted by using maximum a posterior adaption (MAP) process to create the GMM language model. In this adaptation process only means μ_i of Gaussian components are adapted. That means all the GMMs have same covariance Σ_i and differ only in means. As a result of this each language model is represented as a concatenation of all GMM Gaussians means vectors[6]. This concatenation of all GMMs Gaussians mean vectors is referred as a GMM-UBM super vectors as shown in the Fig.1. Now these super vectors are given to the SVM for classification (identification) purpose.

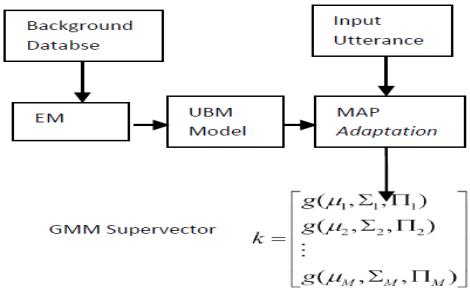

Fig.1: Process of GMM-UBM super vectors

C. GMM-UBM Super Vectors with SVM

SVM is a binary classifier based on a hyper plane separators principle. The basic idea is to map the data into feature space (hyper plane). This feature space is the basis for SVM approach to determine hyper plane or liner decision boundary. This decision boundary optimally separate two classes. The separator is chosen such way that it maximize the distance between the hyper plane and close to the training support vectors. The main objective of SVM is the kernel, which represent an inner product in the SVM high dimensional space (feature space).

The SVM maps the input vector from X to a high dimensional feature space with a mapping function f(x). For the given observation X, the SVM SVM discriminate function having the from

$$f(x) = \sum_{i=1}^{N} \propto_i t_i k(x, x_i) + d$$

In the above equation N indicate the number of support vectors, t_i is the ideal output and α_i represent the weight for the support vector x_i. The $k(x,x_i)$ is the kernel function in the above equation, it is an inner product of two support vector x,x_i in the high dimensional space and it has to satisfy the Mercer condition. The ideal outputs having the values are either 1 or -1, depending the upon whether the corresponding support vector belongs to class 0 or class1. The α_i is obtained through a training process. For the identification (classification), a class decision is based on the value of f(x), is above or below a threshold[7].

III. GMM BASED LID SYSTEM

First from the training data, 13 dimensional MFCC feature vectors are extracted. Using these MFFC feature vectors GMM model are created. Let λ^1 is the GMM language model created from the training data. For each language under consideration one corresponding language model is created. This process is repeated for all the languages under consideration and separate language models are created for each language. For all language are represented by GMM model $\lambda^1, \lambda^2, \ldots \lambda^K$ respectively as explained in the previous section. The X is the sequence of feature of feature vectors representing the test speech utterance. The objective is to identify the language which is maximizing the likelihood of spoke utterance against the set of language models.

IV. GMM-UBM LID SYSTEMBASED ON SVM

The training phase consists of two steps. They are generation of GMM-UBM super vectors and training of these super vectors using SVM. During first step a 13 dimensional MFCC feature vectors are extracted from the set of different language speech data with 20ms frame size and 10ms frame shift. Using these feature vectors language independent GMM-UBM are trained to represent general speech characteristics. Followed by Map adaptation is used to get the GMM model from UBM model. Here only GMM means are adapted. The adapted GMM mean vectors are concatenated and these vectors referred as a GMM-UBM super vectors. In the second step, the SVM is trained using GMM-UBM super vectors by a employing one against all strategy for S class classifier. The trained SVM classifier classify the super vectors as positive examples (labeled as +1) indicate the current language model (one) and negative examples (labeled as -1) indicate the combined remaining language model.

In the identification phase, the test speech data representing the sequence of feature frame vectors X={ $x_1, x_2, \ldots, x_l$} is calculated against SVM classifier. For each frame, the feature vector is classified by SVM classifier. The output of SVM classier for each vector is the index test score of the identified language. The output of each vector are represented in index test score vector.

Finally the index test scores vector of all frames are combined using an averaging step to give an overall utterance score from which the language is identified.

V. EXPERIMENTS

For the LID study OGI_MLT speech corpus is used. The speech corpus comprises of 12 languages. For each language consists of ten speakers of male and female. Present for LID task five languages are considered. The five languages are namely Hindi, Kanda, Telugu, Tamil and English. First GMM LID system is developed. Next the LID system based on GMM-UBM super vectors with SVM classifier performance is compared and analyzed with the conventional GMM based LID system by varying the number of mixture components and test speech duration.

A. Conventional GMM LID System

Here five language GMM LID system is developed as follows. The GMM based LID system is a two stage procedure comprises of training and testing. During the training, each enrolled language is trained using language training speech. Each language is trained using EM algorithm to generate an GMM language model. In the LID testing phase from the test speech utterance MFCC feature vectors are extracted and score is calculated against all the language models using log likelihood. Which language model yields maximum score declared as the identified language.

B. GMM-UBM Super Vector LID System Based on SVM

Here five language GMM-UMB LID system is developed based on SVM as follows. The GMM-UBM based SVM LID system is also consists of two stage stages, such as creation of GMM-UBM super vectors followed by SVM training and testing. First GMM-UBM super vectors are created and followed by SVM modeling. Next using test data the language identification is performed.

VI. RESULT AND DISCUSSION

The five language GMM LID system performance is analyzed with varying number of mixture components and test speech duration is analyzed. For each language LID systems are developed by varying the number of mixture components16 ,32,64 and 128. The average LID performance of three test durations of 2s, 3s and 5s for five languages are shown in the Table.I.

Table.I: The average of five language GMM LID performance for varying length of Gaussian components and test duration.

No of Mixture Components	1Sec	2Sec	3Sec
32	64.58	74.98	78.96
64	72.06	81.50	85.55
128	76.83	84.80	89.80
256	80.76	87.97	94.30

The five language GMM-UBM LID based on SVM system performance is analyzed with varying number of mixture components and test speech duration is analyzed as shown Table.II.

Table.II: The average of five language GMM-UBM super vector based on SVM LID performance for varying length of Gaussian components and test duration.

No of Mixture Components	1Sec	2Sec	3Sec
32	75.21	82.92	90.71
64	80.24	87.54	93.55
128	83.10	88.81	94.98
256	85.75	90.30	95.27

Here the LID performance of conventional GMM system is compared with GMM-UBM super based on SVM system as shown in Table.III. The comparison is performed by increasing the number of mixture components and test speech duration. From the table.3 it is observed that the number of Gaussian mixture components are increased the LID performance is also increased. It is also observed the performance of GMM-UBM system based on SVM is superior when compared to the conventional GMM system.

Table.III: The average performance comparison of five language GMM and GMM-UBM super vector SVM LID system for varying length of Gaussian components and test duration.

No of Mixture Components	1Sec		2Sec		3Sec	
	GMM	GMM-UBM based on SVM	GMM	GMM-UBM based on SVM	GMM	GMM-UBM based on SVM
32	64.58	75.21	74.98	82.92	78.96	90.71
64	72.06	80.24	81.50	87.54	85.55	93.55
128	76.83	83.10	84.80	88.81	89.80	94.98
256	80.76	85.75	87.97	90.30	94.30	95.27

VII. CONCLUSION

This paper explores the robustness of GMM-UBM super vectors for LID task based on SVM. For this purpose a novel approach is introduced for a new hybrid LID system using GMM-UBM super vectors. Here we developed a LID system incorporating GMM-UBM super vectors based on SVM. In conventional GMM based LID system, it require large amount of training data to capture the variations present in the phonotactic-constraints imposed by the language. This can be overcome by using GMM-UBM super vectors for the LID system. The GMM-UBM LID with super vector based on discriminative classifier support vector machine for better classification and good identification performance. Further, a comparison of GVM-SVM LID system performance and GMM-UBM system performance is investigated. The performance analysis proved that the GMM-UBM super vector SVM LID system is superior when compared to GMM LID system. It also observed the LID system performance is increased with increase in the number of Gaussian mixture components.

REFERENCES

1. Campbell, W.M., Sturim, D.E., Reynolds, D.A.: Support vector machines using GMM supervectors for speaker verification, IEEE Signal Processing Letters 13(5), 2006.
2. W. M. Campbell, "A Covariance kernel for SVM language recognition," Int. Conf. Acoust. Speech and Signal Process., 2008.
3. H Li, B Ma, CH Lee, "A vector space modeling approach to spoken language identification". IEEE Transactions on Audio, Speech, and Language Processing. 15(1), 2007.
4. E.Singer et al (2003). "Acoustic, Phonetic, and Discriminative Approaches to Automatic Language Identification", *In Proc. Eurospeech, 2003*.
5. E. Wong and S. Sridharan, "Methods to improve Gaussian mixture model based language identification system," in Proc. Int. Conf. Spoken Language Processing (ICSLP-2002), 2002.
6. W. M. Campbell, D. E. Sturim, D. A. Reynolds, and A.Solomonoff, "SVM based speaker verification using a GMM super vector kernel and NAP variability compensation," Int.Conf. Acoust. Speech and Signal Process, 2006.

7. C. H. You, K. A. Lee and H. Li, "An SVM kernel with GMM-supervector based on the Bhattacharyya distance for speaker recognition," IEEE Signal Processing Letters, vol. 16, no. 1, 2009.

AUTHORS PROFILE

Dr A.Nagesh is currently working as a professor in CSE at MGIT, Hyderabad. He completed B.E and M.Tech from Osmania University, Hyderabad in 1996 and 2002 respectively. He did Ph.D in CSE from JNTUH, Hyderabad in the year 2012. He is having total 22 years of teaching experience. At present he his supervising five Ph.D students. Total he is having 40 publications in national & international journals. His research areas includes pattern recognition , speech processing and data mining.

.