INTERNATIONAL JOURNAL FOR ADVANCE RESEARCH IN ENGINEERING AND TECHNOLOGY
WINGS TO YOUR THOUGHTS.....

Improved Speaker Recognition using Multiple Kernel Learning Algorithms

Aparna Mohan[1], Sulphikar A[2]

[1]M.Tech Scholar, LBSITW, Poojappura, Thiruvananthapuram 695012,
aparnamohan619 @gmail.com
[2]Associate Professor, LBSITW, Poojappura, Thiruvananthapuram 695012
sulphis@gmail.com

***Abstract**: Support Vector Machine (SVM) is the popularly used classifier technique in speaker recognition as it is more robust and has got better generalization performance. Moreover it has the ability to classify the unseen data accurately. But in SVM, selection of appropriate kernel is a question for efficient implementation. Therefore Multiple Kernel Learning (MKL) is proposed. Multiple Kernel Learning is the set of machine learning methods that use a predefined set of kernels and learn an optimal combination of kernels as part of the algorithm. Using a specific kernel as a source of bias, and allowing a learner to choose among set of kernels, a better solution can be found. A performance analysis and comparison of the proposed method MKL with previous method SVM with a single kernel is doing here .The accuracy of the proposed method is increased when compared to the previous method.*

***Keywords:** Mel Frequency Cepstral Coefficients, Multiple Kernel Learning, Support Vector Machine*

1. INTRODUCTION

Speaker recognition is a process that verify a person's claim of his/her identity using the features extracted from their voice. The speech signal carries information about the speaker, the message to be conveyed, language, emotion and so on [1]. Speaker recognition has two phases. The enrollment phase and the verification phase. In the enrollment phase a number of features will be extracted after recording the speaker's voice and that forms a template, a model or a voice print. The verification phase compares an utterance of speech sample with a previously created voice print. Speaker recognition can be done as either text independent or text dependent recognition. In text independent speaker recognition the text spoken by the speaker is unknown to the system. Here different text is used for enrollment and verification. In text dependent speaker recognition the text spoken by the person is already known to the system. Here the same speech will be used for enrollment and the verification phase [2]. Speaker recognition can be classified as speaker identification and speaker verification. Speaker identification identifies the person who is talking from a given set of known voices of speakers [14]. Here a 1: N match is found where N represents the number of templates in the database. Speaker verification is the verification for the identity of a person of his claim. In speaker verification there will be only a 1:1 match which is between a speaker's voice and a template [2].

The anatomical structure of vocal tract differs from person to person. That is it is unique for every person and hence every person has different speech. That is why the speakers can be identified using the information available in the speech. Recognizing the speaker by his/her voice is termed as speaker recognition. Selection of suitable features along with methods to extract them is known as feature selection and feature extraction. These extracted features forms a test template. During classification the test template is compared with a reference template and a similarity measurement is computed between them. If the measurement is within a threshold then the identity claim is accepted else rejected. Based on threshold value two results are possible. One is False acceptances (false claim is accepted) and other is False rejection (owner of the identity is rejected). So usually the threshold is set high. But this can lead to many false rejections that is undesirable. Keeping threshold low while not giving any false rejections may result in many false acceptance. Hence in some classification the threshold will have two values, low and high. If the measurement is below the low value then the identity claim is accepted and if it is above the high value then the claim is rejected. If it is between the low and high values then further classification is done [4].

Several techniques are used to perform feature extraction. The most popularly used methods are Mel Frequency Cepstral Coefficient (MFCC), Linear Predictive Coding (LPC), and Perceptual Linear Prediction (PLP). Mel Frequency Cepstral Coefficient (MFCC) [5]-[11] is widely used in speaker recognition as it can extract both linear and nonlinear features. Thus it can be used to extract dynamic features as well. MFCC is based on human auditory system [13]. Human perception of sounds does not follow a linear scale which is above 1 kHz. In MFCC the frequency scales are placed on a linear scale for frequencies below 1 kHz and on a log scale for frequencies above 1 kHz [6]. It contain both time and frequency information of the signal and thus it is more useful for feature extraction [3]. Linear predictive coding (LPC) can be defined as a digital method which encodes an analog signal by predicting a particular value as a linear function of the past values of the signal [7]. Human speech is produced in the vocal tract, which can be approximated as a variable diameter tube. Linear predictive coding is an approximation of the vocal tract that is represented by this variable diameter tube. The important aspect behind LPC is that it allows the prediction of the next sample by a linear combination of previous samples [6]. Perceptual Linear Prediction (PLP) [6] describe the psychophysics of human hearing more precisely in the feature extraction process. In order to derive an estimate of the auditory system PLP uses three concepts from the psychophysics of hearing. They are, the critical-band spectral

INTERNATIONAL JOURNAL FOR ADVANCE RESEARCH IN ENGINEERING AND TECHNOLOGY
WINGS TO YOUR THOUGHTS…..

resolution, the equal-loudness curve, and the intensity-loudness power law. An autoregressive all pole model will then approximate the auditory system. When compared to LP analysis, PLP analysis is more consistent with human hearing and it is computationally more efficient. It gives a low dimensional representation of speech [8].

There are some commonly used classification algorithms for speaker recognition. The simplest one among them is Vector Quantization (VQ) [9] which is otherwise called centroid model. It was introduced in 1980s.Vector quantization is the process that take large number of feature vectors of a particular speaker and produce smaller set of feature vectors. These feature vectors forms the centroid of distribution. Centroid is the point spaced so that there is only minimum average distance to every other point. During training a speaker specific codebook is generated for each known speaker by clustering the training feature vectors. During the recognition phase when an input utterance of an unknown speech comes it will be vector quantized using each trained codebook. That is the total VQ distortion will be computed. The speaker corresponding to the VQ codebook with minimum total distortion will be identified as the speaker of the unknown utterance. Gaussian Mixture Model (GMM) is one of the commonly used classifier and it is a density estimator [6]. A Gaussian mixture density can be considered as the weighted sum of M component densities. Where x is a D-dimensional random vector, p_i i=1, 2...M are the mixture weights, $b_i(x)$ where i=1, 2... M are the component densities. In this model μ_i , $\sum_i$ represents the mean and covariance of the i^{th} mixture respectively [10] During training when the data x_1 , x_2 , x_3 ...., x_n and the number of mixtures M are given then the μ_i , $\sum_i$, p_i is learned using expectation maximization. During recognition for an unknown utterance set of feature vectors will be extracted that is x_1 , x_2 , x_3 ...., x_l and distance of the given sequence from the model is obtained by computing the log likely hood of the given sequence of feature vectors when the data is given. The speaker will be corresponding to the model with the highest score [7].

Support Vector Machine (SVM) is a powerful discriminative classifier which is popularly used today than other classifiers. SVM can be used with prosodic, spectral and high level features. SVM is a binary classifier [15] which maps the given input to a high dimensional plane and it separate the classes with a hyper plane. Currently this is considered as a robust method for speaker verification .this method has got more generalization performance as it has the ability to classify the unseen data accurately [12]. SVM is a linear learning machine that is expressed in a dual fashion. That is data appear only in the inner product. SVM basically operate in a feature space induced by a kernel.The main objective of SVM is to maximize the margin. The hyperplane that maximizes the margin should be considered. SVM is currently used in most of the speaker recognition applications for the following reasons. Training is comparatively easy. It scales well to high dimensional data. Tradeoff between classifier complexity and error can be controlled explicitly. Apart from these advantages SVM has got a big challenge [16]. This is regarding the choice of kernel. SVM allows nonlinear classification by mapping the data points to a high dimensional space. Due to the computational overhead of mapping to a high dimensional space SVM allows to perform an implicit mapping using Kernel function. But the kernel should be chosen in such a way that the performance of the system can be increased. But the choice of appropriate kernel is an issue in SVM. As a solution to this problem a new method is proposed which is Multiple Kernel Learning.

Multiple Kernel Learning [16] is the set of machine learning methods that use a predefined set of kernels and learn an optimal combination of kernels as part of the algorithm. Using a specific kernel as a source of bias, and allowing a learner to choose among set of kernels, a better solution can be found. In this paper a performance analysis and comparison of the proposed method MKL with the previous method SVM with a single kernel is doing and results are compared.

The following sections include the details of proposed system. Section 2 describes the proposed system architecture and multiple kernel learning algorithms. Results obtained from the work is specified in section 3.Section 4 concludes the work and describes the future work.

2. PROPOSED SYSTEM

Success of SVM is dependent on the choice of good kernel which is typically hand-crafted and fixed in advance. The aim of this work is to develop a learning method which helps to choose a kernel or to use a combination of them and to improve the performance. The method learns kernel from training data. It focuses on how the kernel can be learnt as a linear combination of given base kernels. The main goal of the proposed method is to learn a classifier and kernel weights.

2.1 System Architecture

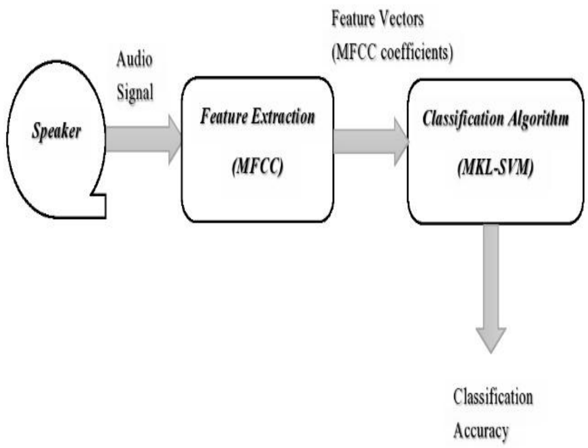

Figure 1: System Architecture

The system architecture of the proposed system is shown in figure above. It consists of mainly two components. Feature extraction and Classification. Feature extraction is one of the important process in speaker recognition. This is the component which is responsible for separating one speech from another. Every speech will be having different individual characteristics embedded in utterance. Feature extraction helps to extract these characteristics. The speech signal /audio signal from the speaker is given to the feature extraction module. The extracted features representing the characteristics of the speech signal will be given to the classification algorithm. In this work Mel Frequency Cepstral Coefficient (MFCC) is used for feature extraction. MFCCs are given as input to the classification algorithm. Classification algorithms classify a new data into one category or the other. Classification can be either supervised or unsupervised. In supervised classification the training samples and its corresponding observations are already known. By using this information we should predict the class of an unknown speech. But in unsupervised classification we does not have any idea about the observations of the training data. This work deal with supervised speaker recognition task and for this Multiple Kernel Learning is proposed.

2.2 Multiple Kernel Learning (MKL)

Multiple Kernel Learning (MKL) is a set of machine learning methods that use a predefined set of kernels and learn an optimal combination of kernels as part of the algorithm. Instead of using one specific kernel function, multiple kernels and its corresponding parameters are used .MKL is used because of the following reasons: (a) Different Kernels will be having different notions of similarity. When we use single kernel we need to find out which works best. Instead of that in MKL it provides a learning method which will find out which works best or will help in using a combination of them. Usually using a single kernel will result in bias. A better solution can be obtained if a learner is allowed to choose the best among them. (b) Different kernels will use inputs from different sources. Thus combining kernel will allow to combine multiple information sources. In MKL the kernel is defined as [16]:

$$\mathrm{K}(x_i, x_j) = \sum_{m=1}^{p} \mathrm{d}_m k_m(x_i, x_j) \qquad (1)$$

where d_m>=0 and k_m are base kernels. The final equation of MKL then becomes:

$$\sum \alpha_i \text{-}(1/2) \sum_i \sum_j \alpha_i \alpha_j \, y_i y_j \sum_{m=1}^{p} \mathrm{d}_m k_m(x_i, x_j) \quad (2)$$

Learning both α_i and the kernel weights d_m in a single optimization problem is called multiple kernel learning problem. In MKL the user only need to specify a set of base kernels. A learning algorithm will then find the combination of these base kernels that is appropriate for the problem at hand. In this direction there are two main lines of work [18]. The first one learns the kernel weights and the parameters of the classifier in a single optimization problem whereas the second line of work uses a two stage approach. SimpleMKL comes under this category. The Twostage approach states that first learn a good combination of the base kernels using the training data and then use the learned kernel to obtain the classifier. Twostage MKL is used for this type of work. In order to efficiently implement MKL the following algorithms are used.

2.3 SimpleMKL

It is a simple Multiple Kernel Learning algorithm which is based on gradient descent of SVM objective value[17]. Given M kernel functions k_1 , k_2 , k_3 …., k_M well suited for a given problem we need to find an positive combination of these kernels such that the resulting kernel is optimal. The kernel function is as follows:

$$\mathrm{K}(x_i, x_j) = \sum_{m=1}^{M} \mathrm{d}_m k_m(x_i, x_j) \qquad (3)$$

$\mathrm{d}_m \geq 0$ and $\sum_m d_m = 1$. For the implementation of this algorithm we need to learn the kernel coefficients d_m and SVM parameters. The algorithms is as follows: Initially we set the kernel weights d_m =1/M for m=1, 2,…M , where M is the number of kernels. The objective value of the dual problem is given by

$$\mathrm{J(d)} = \sum \alpha_i \text{-} (1/2) \sum_i \sum_j \alpha_i \alpha_j \, y_i y_j K(x_i, x_j) \qquad (4)$$

where $K(x_i, x_j) = \sum_{m=1}^{M} \mathrm{d}_m k_m(x_i, x_j)$. We should minimize d. The main goal is to compute the kernel weights and SVM parameters .For that initially keep the value of d_m for m=1, 2...M fixed and then compute the parameters of SVM that is α_i, α_j, b. After finding the SVM parameters, compute the kernel weights. Obtain the new values of kernel weights and update the value of d_m for m=1, 2,… M.Th. following steps should be repeated until the stopping criteria is met.

Compute the objective value J (d) using an SVM solver with $K = \sum_{m=1}^{M} \mathrm{d}_m k_m$. Compute $\frac{\partial J}{\partial \mathrm{d}_m}$ assuming do not dependent on d. That is

$$\frac{\partial J}{\partial \mathrm{d}_m} = (1/2) \sum_i \sum_j \alpha_i \alpha_j \, y_i y_j k_m(x_i, x_j) \qquad (5)$$

for all m. The projected gradient will be taken as the descent direction D. Now the value of d should be updated with the descent direction .This is computed by

$$\mathrm{d}_m{}^{new} = \mathrm{d}_m{}^{old} + \Upsilon D \qquad (6)$$

Where ϒ is the step size. SimpleMKL is faster for small number of training samples and large number of kernels [17].

2.4 SimpleMKL with individual features

The algorithm is same as Simple MKL algorithm. The difference from SimpleMKL is that it does not give the entire feature vector as input to kernels. Rather it create M number of kernels, where M indicate the dimension of the input feature vector. Then each kernel will be given individual features as input. The kernel function will take the index of the feature and will process only the feature corresponding to the index. The constraint here is to create the number of kernels corresponding to the number of dimensions of the input vector. The algorithm perform better than ordinary Simple MKL.

2.5 Twostage MKL

Here the kernel learning problem is considered as the standard linear classification problem in a new instance space. The method says that any linear classifier with weights μ will corresponds to the linear combination of base kernels with weights μ [18] .Thus the problem of finding a good kernel combination reduces to the problem of finding a good linear classifier in the new space. The biggest advantage of two stage MKL is that any binary classification method can be adapted to solve the MKL problem. Consider a classification problem where the instances are drawn from a distribution P over X x Y, where Y is set of discrete labels. Let us assume that we have access to p positive semi definite base kernel functions k_1 , k_2 , k_3 …., k_M where k_i:X xX-> R.

INTERNATIONAL JOURNAL FOR ADVANCE RESEARCH IN ENGINEERING AND TECHNOLOGY
WINGS TO YOUR THOUGHTS…..

The main goal of Twostage MKL is to learn combination of these kernels that itself positive semi definite and is good for the classification task at hand. We have X_1, $X_2, X_3 \ldots X_N$ samples and Y_1 , $Y_2, Y_3 \ldots Y_N$ labels. The samples and lables in new space are shown in table 1.

The new space thus obtained is termed as K- space, the samples in the new space is called K-instance or K-example, and the labels in the new space is called K-labels. Any function h:R^p- > R in this space induces a similarity function K_h between instances in the original space. That is

$$K_h(X_1, X_2) = h\ (k_1(X_1,X_2), k_2(X_1,X_2),.k_p(X_1, X_2)) \quad (7)$$

Table 1: Samples and labels in K- space

Samples in new space	**Labels in new space**
K(X_1,X_1)=(k_1(X_1,X_1),k_2(X_1,X_1),..k_p(X_1,X_1))	$Y_1 . Y_1$
K(X_1,X_2)=(k_1(X_1,X_2),k_2(X_1,X_2),.k_p(X_1, X_2))	$Y_1 . Y_2$
●	●
●	●
K(X_2,X_1)=(k_1(X_2,X_1),k_2(X_2,X_1),..k_p(X_2, X_1))	$Y_2 . Y_1$
●	●
●	●
K(X_N,X_N)=(k_1(X_N,X_N), k_2(X_N,X_N), k_p(X_N X_N)	$Y_N . Y_N$

The new space thus obtained is termed as K- space, the samples in the new space is called K-instance or K-example, and the labels in the new space is called K-labels. Any function h:R^p- > R in this space induces a similarity function K_h between instances in the original space. That is

$$K_h(X_1, X_2) = h\ (k_1(X_1,X_2), k_2(X_1,X_2),.k_p(X_1, X_2)) \quad (7)$$

If K_h is positive semi definite and is a valid kernel we say h is a K-classifier. Consider the example of two base kernels [18]. Each point in the figure is a labeled. K-example corresponding to the pair of original instances. For a linear K-classifier, the value of its induced kernel for a pair of original instances $K_\mu(x_1, x_2)$ is the projection of corresponding K-example on vector μ.

In the figure the left and the centered figure show the case of μ = (0, 1) and (1, 0). Both these cases are suboptimal. The case of μ = (1, 1) is only optimal which is shown in the right figure as it separate the instances of one class from the other. Thus the K-classifier can be considered as an optimal classifier.

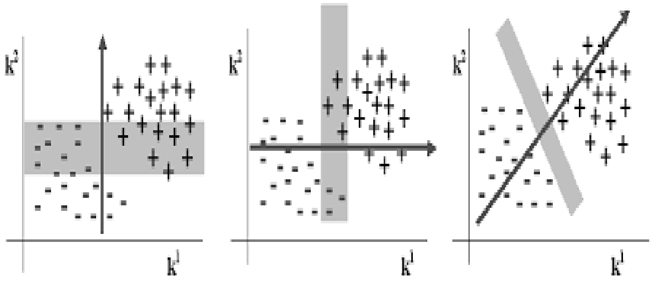

Figure 2: K-space for two base kernels (p=2) [18]

3. RESULTS

The experiment is conducted using two sample datasets, XM2VTS and Telephone based speaker identification dataset from India. In XM2VTS there are 5 speakers each one having 24 audio files and in the Telephone based speaker identification dataset from India there are 20 speakers each having 5 audio files. The experiment starts with a feature extraction module. Mel Frequency Cepstral Coefficient (MFCC) is used for feature extraction. Each audio file is given as input to the feature extraction module. The read speech file will be converted to a vector of sampled data. The sampled speech will be framed to a number of frames. Let N be the number of frames. N varies from one speech to another which depends on its length. Then for each frame further processing is done. The frame duration is taken as 25 ms and the frame shift is taken as 10ms. Now from each frame 13 MFCC coefficients are generated. Thus a single audio signal after feature extraction gives a 13xN vector, where N is the number of frames. Then the frames will be averaged to obtain a 13x1 vector. The result will be extracted feature vector of a speech signal. This 13xN matrix is converted to a 13x1 column vector by finding the average of frame values. This 13x1 column vector is the MFCC coefficient of that audio signal. This is the MFCC of the speech signal. This is repeated for all speech samples of a speaker. Similar processing is applied to all speakers. The XM2VTS dataset contain 120 feature vectors and the Telephone based speaker identification dataset from India contain 100 feature vectors.

The next module of the system is classification. The extracted feature vectors should assign a label corresponding to the class it contain. The XM2VTS dataset contain 5 classes corresponding to 5 speakers which can give labels as 1, 2,3,4,5.

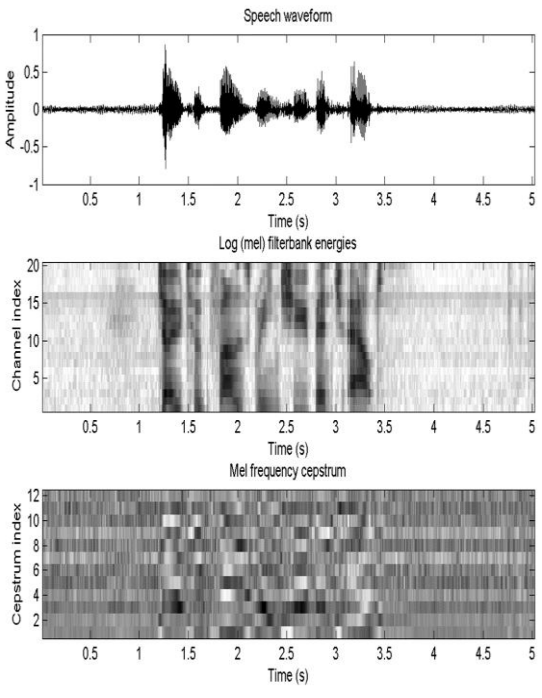

Figure 3: MFCC of a speaker's speech

INTERNATIONAL JOURNAL FOR ADVANCE RESEARCH IN ENGINEERING AND TECHNOLOGY
WINGS TO YOUR THOUGHTS…..

The XM2VTS dataset contain 120 feature vectors and the Telephone based speaker identification dataset from India contain 20 classes corresponding to 20 speakers which can give labels as 1,2,3,4,...20. Then the dataset should be given as input to the classification algorithm. The data set should be divided into training and test samples. In this experiment Leave-one-out cross validation is used for dividing into training and test samples. Where at a time 1 sample will be taken as the test sample and remaining samples will be taken as the training sample. This is repeated for all feature vectors. That is at each iteration a different feature vector will become the test sample and remaining become the training sample. Thus all feature vector will become a test sample at least once. The average accuracy obtained with all feature vectors as test sample will be taken which will be the classification accuracy of the algorithm.
The experiment is applied to the following algorithms. SMOSVM,SimpleMKL , SimpleMKL with individual features, Twostage MKL. This is done with both data sets. The results are shown below.

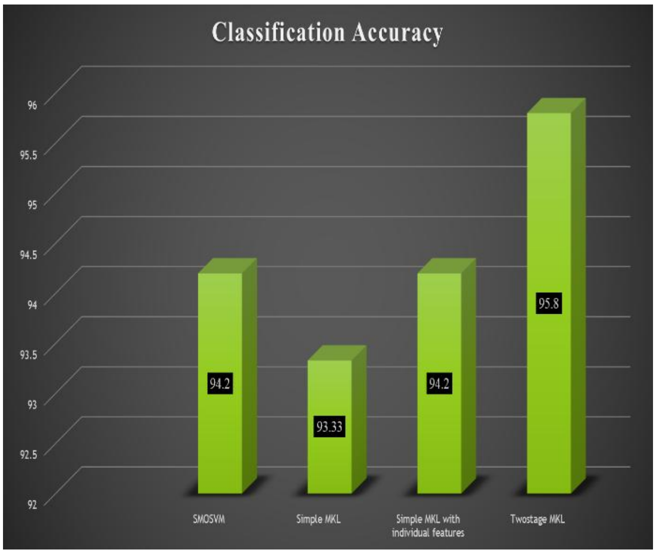

Figure 4: Graph showing the accuracy of four methods using XM2VTS dataset

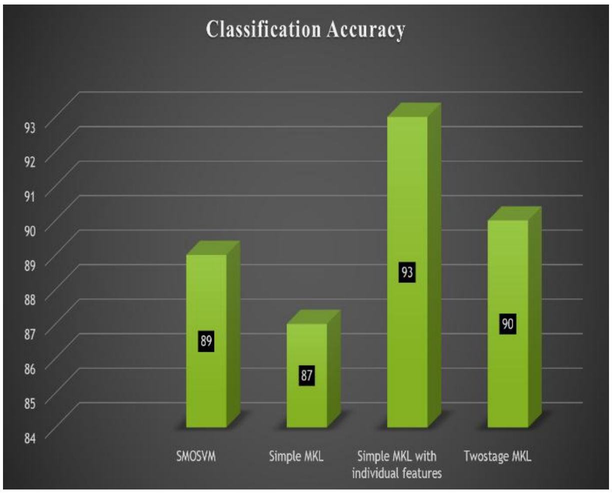

Figure 5: Graph showing the accuracy of four methods using Telephone Based Speaker Identification Dataset

The results shows that the classification accuracy of Multiple Kernel Learning has improved when compared to the SVM algorithm which uses only a single kernel. The experiment compare the different algorithms that use different kernel combinations. The experiment proves that Twostage MKL and SimpleMKL with individual features provide efficient method of combining the kernel functions.

4. CONCLUSIONS AND FUTURE WORK

Support Vector Machine is a discriminative classifier popularly used but there is a question of appropriate kernel to be used. Thus Multiple Kernel Learning is proposed which uses a learning method to choose a kernel or allow to use a combination of them. The different MKL algorithms were tested with two sample datasets. The results shows that the problem of choosing suitable kernel for classification has been solved with Multiple Kernel Learning. The performance accuracy of MKL has increased when compared to SVM with a single kernel. In XM2VTS Two stage MKL gives better classification accuracy than Simple MKL with individual features where as in Telephone based Speaker Identification dataset due to noise constraints and reduced number of training samples than in XM2VTS SimpleMKL with individual features works better than Twostage MKL .From the experiments done it can be concluded that the performance accuracy of TwoStage MKL and SimpleMKL with individual features have increased and had found to be an efficient method for choosing a good kernel combination.
In our future work, features can be combined from multiple feature extraction techniques. This combined features will form the feature vector and will be given as the input to the classification algorithm.

ACKNOWLEDGMENT

I feel unequivocal gratification in acknowledging the service and co-operation rendered by Mr.Sulphikar A., Associate Professor, Department of Computer Science and Engineering, and Mr. Sandeep Chandran, Assistant Professor, Department of Computer Science and Engineering, for their valuable suggestions and guidance throughout the completion of my work.
I also acknowledge with grateful thanks the authors of the references and other literature referred to in this paper.

References

[1] Zia Saquib, Nirmala Salam, Rekha P. Nair, Nipun Pandey, and Akanksha Joshi, "A Survey on Automatic Speaker Recognition Systems", Procedia Computer Science,00(2009)000-000.

[2] Marcos Faundez-Zanuy, Enric Monte-Moreno, "State- Of-the- Art in Speaker Recognition",IEEE A and E Systems Magazine,00(2009)000-000.

[3] Mr.Yoghesh Dawned, Dr. Mukta Dhopeshwarkar, Dr. Babasaheb Ambedkar, "Analysis Of Different Feature Extraction Techniques for Speaker Recognition", International Journal Of Advanced

INTERNATIONAL JOURNAL FOR ADVANCE RESEARCH IN ENGINEERING AND TECHNOLOGY
WINGS TO YOUR THOUGHTS…..

Technology and Engineering Research (IJATER), ISSN2250-3536, vol.5, Issue 1, Jan.2015.

[4] Nilu Singh, R.A .Khan, RajShree, "Applications of Speaker Recognition" International Conference on Modelling, Optimisation and Computing (ICMOC2012), ELSEVIER, Procedia Engineering 38 (2012) 3122-3126.

[5] Nilu Singh, R. A. Khan, Raj Shree," MFCC and Prosodic Feature Extraction Techniques," International Journal of Computer Applications, vol. 54, no.1, Sept.2012.

[6] Jeet Kumar, Om Prakash Prabhakar, Navneet Kumar Sahu , "Comparative Analysis of Different Feature Extraction and Classifier Techniques for Speaker Identification Systems :A review ,"International Journal of Innovative Research in Computer And Communication Engineering , vol.2, issue 1, Jan.2014.

[7] Dr. Mustafa Dhiaa Al-Hassani, Dr. Abdul kareem A Kadhim, "Design A Text Prompt Speaker Recognition System using LPC Derived Features" , 13^{th} International Arab Conference on Information Technology (ACIT'2012),ISSN- 1812-0857,Dec.10-13.

[8] Hynek Hermansky, "Perceptual Linear Predictive Analysis of speech", J.Acoust.Soc.Am.87 (4), April 1990.

[9] Hemlata Eknath Kamale, Dr. R. S. Kawitkar , "Vector Quantization Approach for Speaker Recognition,"International Journal of Computer Technology and Electronics Engineering (IJCTEE), vol.3 ,March-April 2013.

[10] Douglas A. Reynolds and Richard C. Rose, "Robust Text Independent Speaker Identification using Gaussian Mixture Speaker Models," IEEE Trans. On Speech and Audio Processing. Vol.3, No:1, Jan.1995.

[11] Shikha Gupta, Jafreezal Jaafar, Wan Fatimah wan Ahmad and Arpit Bansal, "Feature Extraction using MFCC," Signal & Image Processing: An International Journal (SIPIJ), vol.,no.4, Aug.2013.

[12] W.M.Campbell,J.P.CampbellD.A.Reynolds, E.Singer, P.A.Torres-Carrasquillo "Support Vector Machines for Speaker and Language Recognition ",ELSEVIER, Computer speech and language Aug.2005.

[13] Varsha Singh, Vinay Kumar Jain, Dr. Neeta Tripathi, "A Comparative Study on Feature Extraction Technique for Language Identification", International Journal of Engineering Research and General Science , ISSN 2091-2730, Vol.2, April-May 2014.

[14] Parvati J Chaudhary, Kinjal M Vagadia,"A Review Article on Speaker Recognition with Feature Extraction", International Journal of Emerging Technology and Advance Engineering, ISSN 2250-2459, Vol.5,Issue 2, Feb.2015.

[15] W. M. Campbell, D. E. Sturim , and D. A. Reynolds, "Support Vector Machines using GMM Supervectors for Speaker Verification, "IEEE Signal Processing Letters, Vol.13, no.15, May 2006.

[16] Mehmet Gonen , Ethem Alpaydm, "Multiple Kernel Learning Algorithms", Journal of Machine Learning Research,12(2011) 2211-2268 .

[17] Alain Rakotomamonjy, Francis R Bach, Stephane Canu, Yves Grandvalet "SimpleMKL", Journal of Machine Learning Research , 9(2008) 2491-2521 .

[18] Abhishek Kumar, Alexandru Niculescu-Mizil, Koray Kavukcoglu, Hal Daume, "A Binary Classification Framework for Two Stage Multiple Kernel Learning" International Conference on Machine Learning, 2012.