ISSN 2321-807X
Volume 13 Number10
Journal of Advances in chemistry

Analysis of Speaker Verification System Using Support Vector Machine

P.Shanmugapriya[a *], Y.Venkataramani[b]
[a]Associate Professor, Department of ECE, Saranathan College of Engineering, Venkateswara Nagar, Panjappur,Tiruchirappalli-620012, TamilNadu, India.*-Corresponding author
shanmugapriya-ece@saranathan.ac.in
[b]Dean (R&D) and PG Professor, Department of ECE, Saranathan College of Engineering,Venkateswara Nagar, Panjappur,Tiruchirappalli-620012, TamilNadu, India.
deanrnd44@gmail.com

ABSTRACT

The integration of GMM- super vector and Support Vector Machine (SVM) has become one of most popular strategy in text-independent speaker verification system. This paper describes the application of Fuzzy Support Vector Machine (FSVM) for classification of speakers using GMM-super vectors. Super vectors are formed by stacking the mean vectors of adapted GMMs from UBM using maximum a posteriori (MAP). GMM super vectors characterize speaker's acoustic characteristics which are used for developing a speaker dependent fuzzy SVM model. Introducing fuzzy theory in support vector machine yields better classification accuracy and requires less number of support vectors. Experiments were conducted on 2001 NIST speaker recognition evaluation corpus. Performance of GMM-FSVM based speaker verification system is compared with the conventional GMM-UBM and GMM-SVM based systems. Experimental results indicate that the fuzzy SVM based speaker verification system with GMM super vector achieves better performance to GMM-UBM system.

Keywords

Gaussian Mixture Model, Fuzzy Support Vector Machine, Speaker Verification System

Academic Discipline and Sub-Disciplines

Modeling the Signals and Analysis of modeling techniques.

SUBJECT CLASSIFICATION

Nonlinear classification

TYPE (METHOD/APPROACH)

Experimental

1. INTRODUCTION

Speaker verification is a method to determine whether a person is who he/she claims to be. Identity given by the claimed speaker and his/ her test speech utterance are the two inputs applied to the system. The system will verify whether the test speech utterance correspond to the claimed identity or not.

For text independent speaker verification system, where there is no prior knowledge of what the speaker will say, the successful model for speaker is Gaussian Mixture Model (GMM). GMM models are trained using the standard Expectation Maximization (EM) training algorithm. Universal Background Model (UBM) represents the whole set of target speakers and background speakers, individual model for each target speaker are obtained through Maximum A Posteriori (MAP) adaptation. Recently, the idea of stacking the mean vectors of GMM model to form a GMM mean super vector has become successful in speaker verification using Support vector Machine [6, 7].

Support Vector Machine (SVM) is a two-class classifier based on the principles of structural risk minimization. SVMs perform a non linear mapping from an input space to an SVM expansion space. Linear classification techniques are then applied in this potentially high-dimensional space.

In the early 1990s, SVMs were first proposed by Vapnik [1] as optimal margin classifier. In pattern recognition works [3], SVM had been used for isolated handwritten digit recognition [2], object recognition [4] and speaker identification [5]. Then, in order to combine the advantage of SVM and the state of art technique GMM-UBM for speaker verification system, a new GMM-SVM system was proposed by Campbell W M et.al [6]. In this approach, the

GMM super vector is the input for SVM. The experiments done by Campbell W M et.al [7] using SVM-GMM and NAP variability compensation with 20 female and 20 male speakers has achieved an error rate of 0.4% and average accuracy rate of 95.1% with 22 order MFCCs for the 2004 NIST speaker recognition evaluation corpus.

The main design component in an SVM is feature space. Since inner products induce distance metrics and vice versa, the basic goal in SVM kernel design is to find an appropriate metric in the SVM feature space relevant to the classification problem. A study on the use of MFCC and SVM for text dependent speaker verification is carried out by Shi-Huang Chen et.al [8]. By using discrete events and their probabilities from speech signal to construct super vectors based on Bhattacharyya distance as input for SVM, Kong Aik Lee et.al [9] obtained an Equal Error Rate (EER) of 5.51% and a Decision Cost Function (DCF) of 2.69. The performance of SVM depends on the selection of Kernel functions used to

DOI : 10.24297/jac.v13i10.5839
www.cirworld.com

compute distances among data points. Mostly used kernel functions are polynomial, linear and Gaussian functions. Since these functions do not use the advantage of inherent probability distributions of data, a deterministic kernel based on KL divergence was proposed by Pedro J. Moreno et.al [10]. Another drawback of GMM-super vector-SVM approach in speaker verification found by Wai Mak and Wei Rao[11] is imbalance between the numbers of speaker class utterances and imposter class utterances. They proposed a method of utterance partitioning with acoustic vector resampling to reduce the error rate due to data imbalance problem. Zhao jian et.al [12] was able to obtain an EER of 4.92% and a DCF of 0.0251 by using GMM-SVM with a Nuisance Attribute Projection kernel.

The outline of this paper is as follows: Section 2 describes GMM-UBM, GMM-SVM and GMM-FSVM development and formation of super vector. Section 3 describes the database used, features extracted and parameters of the proposed system. Simulation results are discussed in section 4.

2. GMM-SVM AND GMM-FSVM

2.1 GMM-UBM System

In a GMM-UBM based text-independent speaker verification system, Universal Background Model (UBM) with a large number of Gaussian mixture components is created based on pool of speech data from target and background speakers [17]. Target speaker models are generated from the UBM by MAP adaptation for the individual target speaker's training utterances [18]. Thus the set of parameters mean, covariance and mixture weights represents the model of a speaker $\lambda_i = \left\{\overline{\mu_i}, \Sigma_i, p_i\right\}$.

2.2 GMM Supervector Formation

GMM-UBM is developed for deriving a target speaker's GMM by adapting the parameters of the model[15]. Only the mean vectors of target speaker model are adapted using (1)-(5). Specifically, given an enrollment utterance with acoustic vector sequence $X = \{\overline{x_1}, \overline{x_2}, ...\overline{x_T}\}$, the mean vectors μ_i of the UBM is used to obtain the adapted mean vectors

$$\hat{\mu}_i = \alpha_i E_i(X) + (1-\alpha_i)\mu_i, \quad i = 1,...,M, \quad \text{where} \tag{1}$$

$$\alpha_i = \frac{n_i(X)}{n_i(X)+r}\ , \tag{2}$$

$$n_i(X) = \sum_{t=1}^{T} \Pr(i|x_t), \tag{3}$$

$$E_i(X) = \frac{1}{n_i}\sum_{t=1}^{T} \Pr(i|x_t)x, \tag{4}$$

$$\Pr(i|x_t) = \frac{\lambda_i p_i(x_t)}{\sum_{j=1}^{M} \lambda_i p_i(x_t)}. \tag{5}$$

The adapted means vectors of all M mixtures are combined to produce the GMM super vector

$$\hat{m}_{[1x\,MD]} = [\hat{\mu}_1, \hat{\mu}_2, ..., \hat{\mu}_M]; \tag{6}$$

The GMM super vector can be thought of as a mapping between an utterance and a high dimensional vector.

As stated in [11], the number of super vectors for target speakers and impostors is increased by means of partitioning the utterance into groups. Then, for each group a super vector is created and this is repeated several times by randomly rearranging the sequence of utterance. This method of utterance partitioning provides sufficient number of super vectors for training the SVM.

2.3. Fuzzy Support Vector Machine Classifier

Support Vector Machine is a two class classifier which classifies using the separating hyperplane [13]. The hyperplane is determined by maximizing the distance between the training vectors and the hyperplane. This hyperplane will be the best decision surface if the training set is linearly separable. The data points on the hyperplane are called as support vectors. If the training set is not linearly separable, then the support vectors can be transformed to a high dimensional space (HDS) with a nonlinear transformation. This nonlinear transformation is represented by Kernel function. Kernel function describes inner product in the HDS (named as feature space) which satisfies the Mercer's condition [14] $K(\mathrm{x}_i, \mathrm{x}_j)$ is Kernel function which provides the distance between two data points in feature space, i.e.,

$$K(\mathrm{x}_i, \mathrm{x}_j) = \varphi(\mathrm{x}_i), \varphi(\mathrm{x}_j) \quad (7)$$

where $\varphi(\mathrm{x}_i)$ is the mapping function from input space to feature space. Then the optimal hyper plane is defined by decision function

$$g(x) = \sum_i \alpha_i t_i K(\mathrm{x}_i, \mathrm{x}_j) + b \qquad 1 \le i \le C \quad (8)$$

where C is the number of data points;

t_i is the target values, $t_i \in \{-1,1\}$ which indicates the class label for the training vectors; α_i s are non negative Lagrangian multipliers;

The values of α_i are computed by solving the quadratic programming problem with linear constraints and these are non zero only for support vectors. The main aim of the binary classification in SVM is to search for a linear hyperplane $g(\mathrm{x}) = w^T z + b$ where $z = \varphi(\mathrm{x})$ denote the corresponding feature space vector with a mapping φ from $\Re^n$ to a feature space z ; w and b are weight and bias vectors which describes the hyperplane. Hence the problem can be stated as

$$\begin{aligned} \langle w, \mathrm{x}_\mathrm{i} \rangle + b \ge 1, \ \forall t_i = 1 \\ \langle w, \mathrm{x}_\mathrm{i} \rangle + b \le 1, \ \forall t_i = -1 \end{aligned} \quad (9)$$

Where $\langle w, \mathrm{x}_\mathrm{i} \rangle$ represents the inner product of w and x_i. Finding out the optimum hyperplane for minimum $\|w\|$ and maximum distance between the hyperplane and training set is not easy for non linearly separable data set. But for non separable training set the optimization problem is modified with the slack variable and a penalty term. This primal training problem has a complicated constraint set and limitation on dimensionality of feature space [3].

Thus the optimal hyperplane can be reformulated as

$$g(\mathrm{x}) = w^T z + b = \sum_{i \in Z_n} \alpha_i K(\mathrm{x}_\mathrm{i}, \mathrm{x}_\mathrm{j}) + b \quad (10)$$

where $b = t_i - \sum_{i \in Z_n} \alpha_i K(\mathrm{x}_\mathrm{i}, \mathrm{x}_\mathrm{j}) \ \ \forall j = 0 \le \alpha_j \le \frac{C}{N}$ where C is the regularization parameter. During training of SVM for a target speaker, in (10), α_i , b and support vectors are optimized to produce a model. Among the optimum support vectors, some training vectors may have more importance than others. Fuzzy theory deals with these issues by saying that a training vector x_i belongs 80% to class +1 and 20% to class -1. This may be achieved by associating a fuzzy membership value $0 \le \mu_i \le 1$ with each training pair (x_i, t_i) [20]. This membership value represents that the vector belongs to the class t_i with membership value μ_i and to class $t_{j \ne i}$ with membership value $1 - \mu_i$. Support vectors those have less contribution in the learning process are neglected based on the membership value. Thus the number of support vectors used in the representation of speaker model is reduced and its performance will be improved. A fuzzy SVM has been proposed as extension to standard SVM. Let X be an input space and T be an output space. Each training pattern is given a label t_i from T and a fuzzy membership value $\sigma \le \mu_i \le 1$ with $i = 1,...,l$. Value of σ must be sufficiently small but greater than zero[21] i.e., since the fuzzy membership value μ_i gives information about the percentage of corresponding data point x_i in the class t_i , FSVM requires a parameter ξ_i to measure the error in the SVM [19]. The product $\mu_i \xi_i$ is the measure of error with different weighting. The primal hyper plane problem in FSVM is stated as:

$$\min_{w,b,\xi \in \Re^n} L(w,b,\xi) = \frac{1}{2} w^T w + C \mu^T \xi \quad (11)$$

such that: $t_i (w^T \mathrm{x}_i + b) \ge 1 - \xi_i \quad i = 1,....,l,$ (12)
$\xi_i \ge 0.$

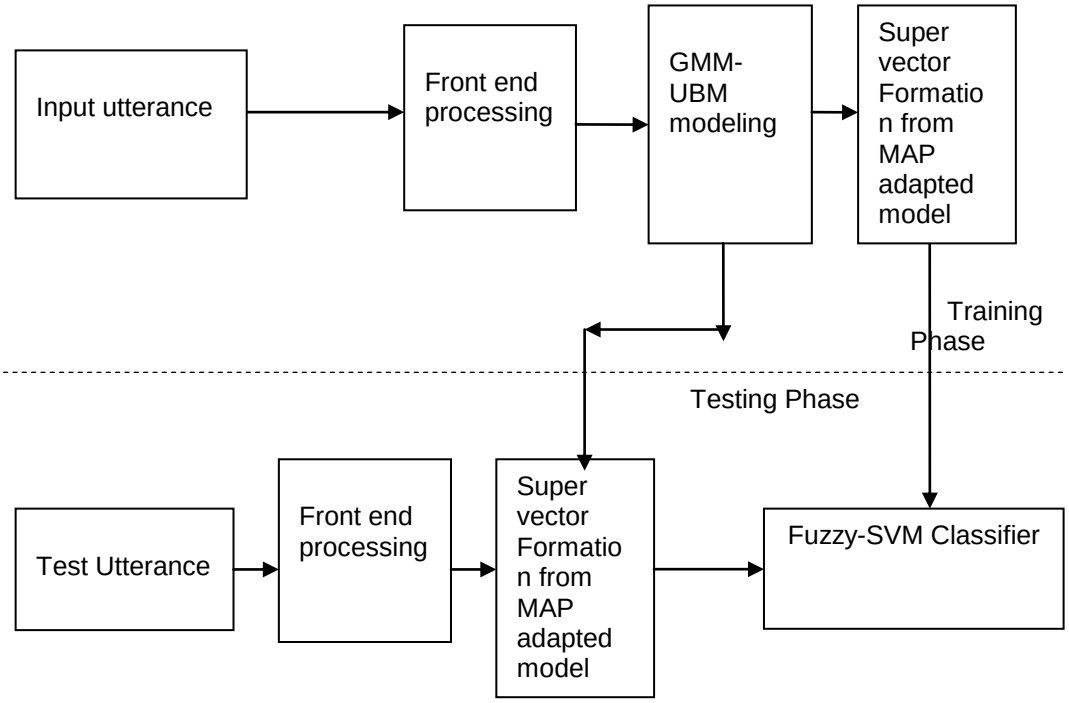

Fig 1. Automatic Speaker Verification system with Fuzzy SVM

where C is a regularization parameter which is used to balance between the minimization of the error function and the maximization of the margin of the optimal hyper plane;

w is the weight vector;

b is the bias;

ξ_i is the slack variable for the data point which is not fitted in the optimal hyper plane. Since the training pair (x_i, t_i) is weighted by the membership value μ_i, the training pair with less μ_i will have less influence in the decision surface than those with larger μ_i value. Solving (11) is a Quadratic Programming problem [22]. This can be solved by transforming the problem into dual problem as

$$\max \; W(\alpha) = \sum_i \alpha_i - \frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j t_i t_j K(\mathrm{x}_i, \mathrm{x}_j) \qquad (13)$$

Subject to $\sum_i t_i \alpha_i = 0, \quad 0 \le \alpha_i \le \mu_i C, \quad i = 1,....,l$

and the Kuhn Tucker conditions are defined as

$$\alpha_i (t_i (w z_i + b) - 1 + \xi_i) = 0, \quad i = 1,...,l. \qquad (14)$$

$$(\mu_i C - \overline{\alpha_i})\overline{\xi}_i = 0, \quad i = 1,...,l. \qquad (15)$$

The process of fuzzy based clustering is based on finding 'k' partitions. Let 'm' be the weighting exponent on each fuzzy membership, and the degree of fuzziness, μ_i be the support vector, and $U = \{u_{it}\}$ where u_{it} is the degree of membership of x_t in the ith cluster and has $0 \le u_{it} \le 1, \; \sum_{i=1}^{k} u_{it} = 1 \quad 1 \le i \le k \;\; and \;\; 1 \le t \le n$.

The dissimilarity function with a A norm distance measure between object x_t and cluster centre μ_i is

$$d_{it}^2 = \|x_t - \mu_i\|_A^2 = (x_t - \mu_i)^T A (x_t - \mu_i) \qquad (16)$$

The new cluster centers are determined such that the dissimilarity function gets minimized. Hence the squared error objective function is

$$J_m(U,\mu,x)=\sum_{t=1}^{n}\sum_{i=1}^{k}u_{it}^{m}d_{it}^{2} \quad (17)$$

Achieving the minimum objective function depends on the change of fuzzy memberships and norm distances with the new cluster centre.

2.4 GMM-Supervectors in Fuzzy SVM

In GMM-FSVM as shown in Fig.1, when the super vector of test utterance is given as the input to FSVM of target speaker, the verification score of the claimed speaker is given by

$$score^{c}=\alpha^{c}K(\mathrm{x}^{c},\mathrm{x}^{t})-\sum_{i\in S^{b}}\alpha_{i}^{b}K(\mathrm{x}^{c},\mathrm{x}^{b_i}) \quad (18)$$

where α^{c},α^{b} are Lagrangian multipliers of claimed speaker and background speakers respectively;

$\mathrm{x}^{c},\mathrm{x}^{t},\mathrm{x}^{b_i}$ are claimed speaker's super vector(test super vector), target speaker's super vector and background speakers' super vectors respectively;

$K(\cdot,\cdot)$ is a kind of distance measure between two super vectors in the high dimensional feature space.

Experiments have been performed for GMM-SVM based system with various kernel functions such as, Linear, Polynomial, Radial, Quadratic and RBF. Polynomial kernel is given by

$$K(x^{c},x^{t})=(x^{c^{T}}x^{t}+1)^{n} \quad (19)$$

where n is the polynomial order.

Radial Kernel is given by

$$K(x^{c},x^{t})=e^{-\frac{1}{2}\left\|\frac{x^{c}-x^{t}}{\sigma}\right\|^{2}} \quad (20)$$

where σ is the width of the radial basis function.

Quadratic kernel is given by

$$K(x^{c},x^{t})=(x^{c^{T}}x^{t}+1)^{2} \quad (21)$$

An attractive feature of the SVM [23] is that the selection of sub clusters is implicit, with each support vectors contributing one local Gaussian functions, and centered at that data point. By the kernel function, the super vectors are positioned on the surface of hyper sphere in the feature space. Then $K(x_i,x_j)=\varphi(x_i),\varphi(x_j)$ is the cosine of the angle between $\varphi(x_i)$ and $\varphi(x_j)$. An FSVM is trained for each target speaker using the GMM super vector of the speaker's enrolment utterances as positive samples, and GMM super vectors of all utterances from background speakers as negative samples.

2.5 Fisher Linear Discriminant Analysis (FLDA)

Since the dimension of the super vector is high, the performance of SVM may get over trained. To avoid this, the system requires dimensionality reduction before using the super vector in the training of FSVM. The technique used for dimensionality reduction which uses label information in finding informative projections from the data is Fisher-LDA. It maximizes the objective which involves the "within class scatter matrix" and the "between classes scatter matrix".the objective function J(w) is given by

$$J(w)=\frac{w^{T}S_{B}w}{w^{T}S_{w}w} \quad (22)$$

where S_B is the "between classes scatter matrix" and S_W is the "within classes scatter matrix". As with eigenspace projection, training super vectors are projected into a subspace. The test super vectors are projected into the same subspace and identified using a similarity measure. Since the scatter matrices are proportional to the covariance matrices, S_B and S_W are calculated from covariance matrices. Initially, for the ith class the scatter matrix Si is calculated as

$$S_{i}=\sum_{x\in X_{i}}(x-m_{i})(x-m_{i})^{T} \quad (23)$$

where m_i is the mean of data in that class. The within class scatter matrix measures the amount of scatter between items in the same class. It is the sum of all the scatter matrices of all classes.

$$S_W = \sum_{i=1}^{C} S_i \quad (24)$$ where C is the number of classes.

The between class scatter matrix measures the amount of scatter between classes. It is calculated as the sum of the covariance matrices of the difference between the total mean and the mean of each class.

$$S_B = \sum_{i=1}^{C} n_i (m_i - m)(m_i - m)^T \quad (25)$$

where

n_i - the number of super vectors in the class i,

m_i - the mean of the super vectors in that class

m - the mean of all the super vectors.

The eigen vectors and eigen values are computed by solving the eigen value problem as

$$S_B V = \Lambda S_W V \quad (26)$$

The Eigen vectors are sorted based on the Eigen values in descending order and the first (C-1) vectors are used as Fisher's basis vectors. The projected super vectors with fewer dimensions can be obtained by projecting all the super vectors on this Fisher's basis vectors.

3. SYSTEM DESCRIPTION

3.1 Database

Speaker verification experiments were conducted using the NIST 2001 SRE database. The NIST 2001 SRE development database consists of 38 male speakers and 22 female speakers. The evaluation database comprises 74 male speakers and 100 female speakers for training, 850 male speakers and 1188 female speakers for testing. The training utterance for each speaker was for 2 minutes and the testing segment duration was less than 60 seconds. Development database is used for model development and initial validation whereas the evaluation database is used for final validation. Equal Error Rate (EER) and the minimum Decision Cost Function (minDCF) are used as metrics for performance evaluation.

3.2 Feature extraction and Feature warping

In this work, we extracted 13-dimensional Mel frequency Cepstral Coefficients from speech signal for 30ms duration with 20ms overlapping. Cepstral Mean Subtraction (CMS) [24] and RelAtive SpecTrAl (RASTA) filtering [25] are two of the standard feature-based channel compensation techniques. But even after CMS and RASTA filtering, channel and handset mismatch can still cause lots of errors. Hence, with CMS and RASTA, recently introduced feature warping technique called Gaussianization [26] is also used to transform the distribution of a cepstral coefficient feature stream to a Normal distribution based on Cumulative Distribution Function (CDF). It is shown that this technique brought significant improvements in recognition rate of the system compared to system based on standard techniques. Then first order and second order deltas are appended to the Gaussianized cepstral vector. The size of the feature vector is now 39.

3.3 GMM-UBM system

The conventional GMM-UBM is used as the baseline system[16] for comparison with GMM-SVM and GMM-FSVM. It is the prerequisite for developing the GMM-SVM system. Gender dependent UBMs were developed using development database and evaluation database of NIST 2001 SRE corpus. Performance comparison is made between three different number of mixture components of UBM. Table I summarizes the number of speakers used for development and evaluation of the system.

3.4 GMM-SVM and GMM-FSVM system

In GMM-SVM & GMM-Fuzzy SVM (FSVM) based systems, the super vectors are formed from the mean vectors of MAP adapted GMMs. In super vector formation, if an entire utterance is used to develop a speaker specific model through MAP adaptation, it yields only one super vector per speaker. But it is not enough for training and testing the

Table 1 Usage of NIST 2001 corpora in the development and evaluation of the system

	Development			Evaluation		
	Target	Background	Impostor	Target	Background	Impostor
Female	17 from devtest train	17 target +80 from evaltest train	5from devtest train and test	80 from evaltest train	80 target + 22from devtest train	20 from evatest train and test
Male	30 from devtest train	30 target +65 from evaltest train	8 from devtest train and test	65 from evaltest train	65 target + 38 from devtest train	9 from evaltest train and test

SVM based speaker model. Hence in this work, Utterance partitioning method proposed by Man Wai Mak andWei Rao[11] is followed. Training utterances of target and background speakers are partitioned into five groups and a super vector is formed for each group. That means five super vectors per speaker (includes target speaker (17) and background speakers (80) in the case of development) are obtained by dividing the enrollment utterance into five sub utterances. These super vectors (97x5=487) are used as negative training vectors for FSVM. The positive training vectors (for target speakers) are obtained by dividing the training utterance into 10 sub utterances. For each sub utterance a super vector is formed from the GMM trained with these sub utterances. This is repeated 20 times by randomly rearranging the sequence of utterance and every time dividing it into 10 groups. Thus for every target speaker there will be 200 super vectors plus one super vector for full utterance. FSVM is trained for each target speaker using 201 super vectors as positive class data and 487 super vectors as negative class data. The same procedure is followed for SVM based system also. Impostor's super vectors are formed by dividing the training utterance of each imposter into 10 sub utterances and one super vector from full utterance. Similarly for testing the target speaker, the test utterance is subdivided into 10 groups and a super vector for each group plus one full utterance super vector are obtained. So, totally, 11 super vectors are used for testing each target speaker.

The reduction in the dimension of super vectors is performed by Fisher's Linear Discriminent Analysis (FLDA). In FLDA, the super vectors are transformed from high dimensional feature space into low dimensional feature space; also it provides the principal directions of channel variability. The Eigen space is generated using both target and background speakers super vectors. Then target, background, impostor and test super vectors are projected into this Eigen space. The projected super vectors are used for training and testing the SVM and also FSVM.

4. RESULTS AND DISCUSSIONS

Experimental results are obtained for baseline GMM-UBM system, GMM-SVM and proposed GMM-FSVM. The performance of the systems is compared by varying the number of mixtures in UBM, MFCC features with and without RASTA filtering and Gaussianization. Also comparison is made between the GMM-SVM systems for various kernel functions. In GMM-FSVM based system the performance analysis is made with various fuzzy membership functions. The results are assessed using Equal Error Rate (EER) and the minimum Decision Cost Function (DCF), defined by NIST as DCF = 0.1*p*miss + 0.99*p*false_alarm. In addition to these parameters, DET curve is also employed as an overall performance criterion.

Table 2 shows the EER and minDCF values of GMM-UBM system with three different numbers of mixtures. The result suggests that GMM-UBM with 256 mixtures performs better than the other two and therefore the number of mixtures was set to 256 for GMM-UBM and from that the super vectors, are obtained for GMM-SVM and GMM-FSVM. The GMMs of target speakers were adapted from the UBM using MAP adaptation [18] with relevance factor set to 16.

Table 2 Performance of GMM-UBM for different Number of Gaussian Components

Method	Mixtures	MinDCF	EER (%)
GMM-UBM baseline system	128	0.2402	22.9
	256	**0.1938**	**21.81**
	512	0.2267	29.34

Table 3 shows the Performance comparison of GMM-UBM and GMM-SVM using MFCC features with and without RASTA filtering and Gaussianization. It shows that the system performs better for MFCC with RASTA filtering and Gaussianized features. Compared to the baseline system with CMS and RASTA, around 20% relative improvement in both EER and

minimum DCF is obtained on NIST 2001 cellular phone data evaluation. Fig.2 shows the effect of RASTA filtering and Gaussianisation in the MFCC features used for GMM-UBM systems.

Table 3 Performance comparisons between GMM-UBM and GMM-SVM with MFCC, MFCC with RASTA filtering and MFCC with Gaussianisation

Model	Min. DCF	EER (%)
Gaussian Mixture Model		
1.GMM-UBM-MFCC	0.1915	21.77
2.GMM-UBM-MFCC-RASTA	0.1745	18.35
3.GMM-UBM-MFCC-RASTA+GAUS	**0.1255**	**13.53**
GMM-Support vector Machine		
1. GMM-SVM-MFCC	0.1180	11.91
2.GMM-SVM-MFCC-RASTA	0.0951	10.01
3.GMM-SVM-MFCC-RASTA+GAUS	**0.0923**	**9.11**

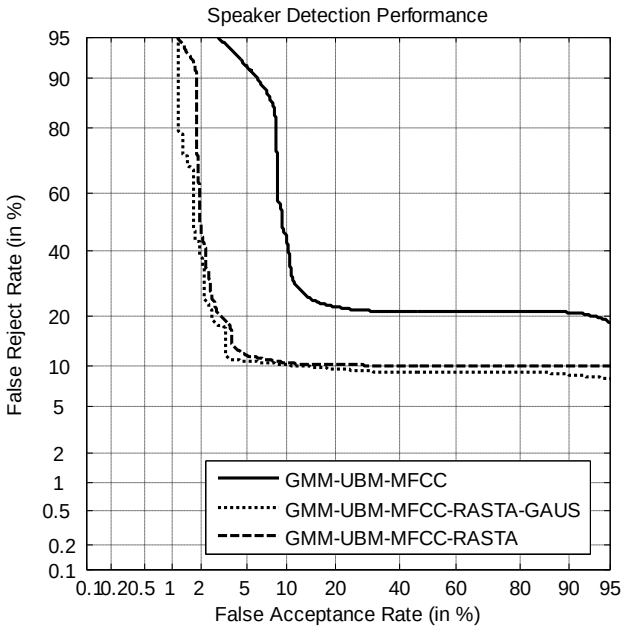

Fig.2.Effect of RASTA filtering and Gaussianisation added in the MFCC features for GMM-UBM system

After the selection of number of Gaussian components in GMM-UBM and the feature warping technique experimentally, the super vectors are formed from the mean vectors of MAP adapted Gaussians. Then, projection of super vectors into the Fisher's Linear Discriminent space yields low dimension discriminent features. Fig.3 shows DET curves for GMM-SVM based system for various Kernel functions. Evidently, the radial kernel function performs better than other kernel functions.

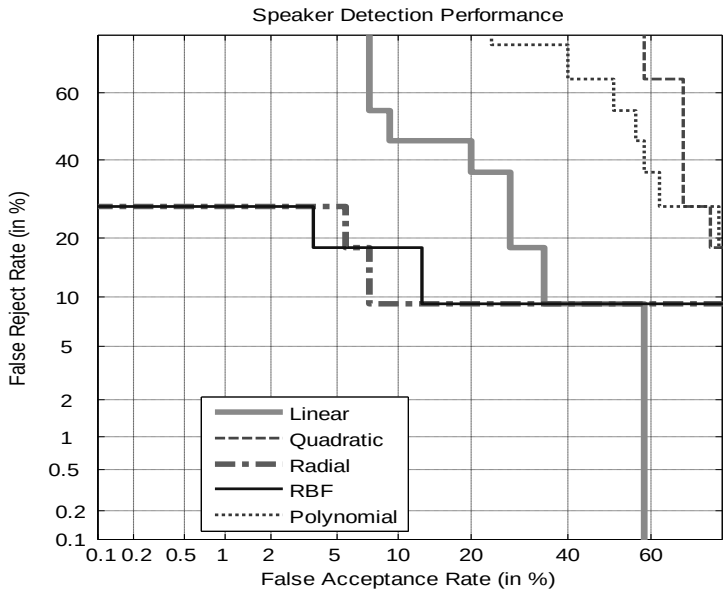

Fig.3 Effect of various kernel functions in GMM-SVM based system.

Performance comparison in terms of EER and minDCF for the GMM-SVM system with various kernel functions are listed in Table 4. The result shows that the radial kernel function in SVM performs better than the conventional linear kernel function. Table 5 shows the performance of GMM-SVM with and without FLDA. There is reduction in the EER of GMM-SVM system with FLDA. This experiment is performed for the gaussianised MF-RASTA features based GMM super vectors and radial kernel function. The result demonstrates the merit of the FLDA in GMM-SVM based speaker verification system.

Table 4 Performance of GMM_SVM for various kernel functions.

Model	Min. DCF	EER(%)	Kernel function
Feature: MFCC+RASTA+GAUS			
GMM-Support vector Machine-with FLDA	0.0923	9.11	Linear
	0.4273	45.65	Quadratic
	0.0818	**8.91**	**Radial**
	0.1091	10.91	RBF
	0.4273	45.56	Polynomial (n=3)

Table 5 Performance of GMM-SVM

Model	Min. DCF	EER(%)
Feature: MFCC+RASTA+GAUS	**Kernel: Radial**	
GMM-Support vector Machine		
1.GMM-SVM	0.0923	9.11
2.GMM-SVM with FLDA	**0.0818**	**8.91**

ISSN 2321-807X
Volume 13 Number10
Journal of Advances in chemistry

Table 6 shows the performance of GMM-FSVM based system with various membership functions. The result shows that the GMM-FSVM system has a significant reduction in EER and minDCF for the radial kernel function and probabilistic sigma membership function.

Table 6 Performance of GMM-FSVM for various membership functions

Model	Min. DCF	EER (%)	Kernel function	Membership Function
Feature: MFCC+RASTA+GAUS				
GMM-Fuzzy Support vector Machine-with FLDA	0.0909	9.56	**Radial**	gaussmf
	0.0812	8.56		gbellmf
	0.081	7.43		gauss2mf
	0.053	**5.02**		**psigmf**
	0.0914	8.52		sigmamf

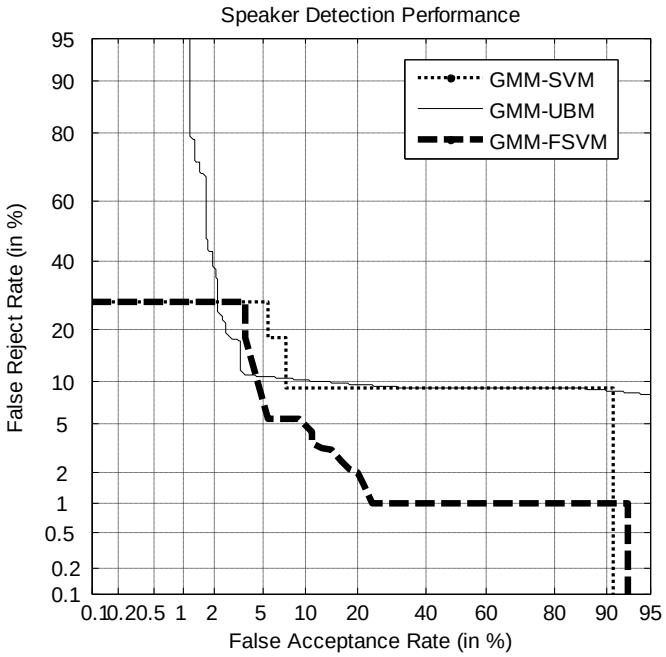

Fig.4 Performance comparison of GMM-FSVM based system with GMM_UBM and GMM_SVM.

Based on the experimental analysis, the proposed method of speaker verification system based on GMM-FSVM performs better than the GMM-UBM and GMM-SVM based systems. The result shown in Fig. 4 in terms of DET curve and EER (5.02%) and minDCF(0.0530) in Table 7 highlights the performance gain that can be achieved by GMM-FSVM.

Table 7 Performance Comparison of GMM-FSVM with GMM-UBM and GMM-SVM

METHOD	EER	DCF
GMM-UBM	13.53%	0.1255
GMM-SVM	8.91%	0.0818
GMM-FSVM	**5.02%**	**0.0530**

CONCLUSION

The commonly used GMM-UBM system can achieve an excellent performance, with EER 13% and DCF 0.12 with gaussianised MF-RASTA features. The SVM system can achieve a comparable performance, with EER 8.9% and DCF 0.082. When introducing the fuzzy membership values of data for training SVM system using Radial kernel in SVM, it leads to the improvement of system EER from 8.9% to 5.02%.

REFERENCES

[1] Vapnik.V.N., "Statistical Learning Theory", New York, USA: Wiley, 1998. www.mit.edu/~6.454/www_spring_2001/emin/slt.pdf
[2] Cortes.C and Vapnik.V, "Support vector networks", Machine Learning, 20, pp.273–297, 1995. Burges.C.J.C., "A tutorial on support vector machines for pattern recognition", Data Min. Knowl. Disc. 2 (2), 121–167, 1998. research.microsoft.com/en-us/um/people/cburges/papers/SVMTutorial.pdf
[3] Blanz.V, Scholkopf.B, Bulthoff.H, Burges.C, Vapnik.V and Vetter.T, "Comparison of view–based object recognition algorithms using realistic 3D models", Artificial Neural Networks—ICANN'96, pp. 251- 256, Berlin, 1996. http://dl.acm.org/citation.cfm?id=684894
[4] Schmidt.M, "Identifying speaker with support vector networks", In Interface '96 Proceedings, Sydney, 1996.
[5] Campbell.W.M, Sturim.D.E, Reynolds.D.A, "Support Vector Machines using GMM Supervectors for Speaker Verification", IEEE Signal Processing Letters, 2006, 13, pp.308-311. https://www.researchgate.net/publication/3343440_Support_vector_machines_using_GMM_supervectors_for_speaker_verification
[6] Campbell.W.M, Sturim.D.E, Reynolds.D.A and Solomonoff. A, "SVM based Speaker Verification using a GMM Super vector Kernel and NAP Variability Compensation", in Proceedings International Conference Acoustics, Speech, and Signal Processing, France, 2006, pp.97-100. https://www.ll.mit.edu/mission/cybersec/publications/.../060514_CampbellW.pdf
[7] Shi-Huang Chen, Yu-Ren Luo, "Speaker verification using MFCC and Support vector Machine", Proceedings of the International Multiconference of Engineers and computer scientists, Hong Kong, vol. 1, March 18-20, 2009. www.iaeng.org/publication/IMECS2009/IMECS2009_pp532-535.pdf
[8] Kong Aik Lee,Chang Huai You, Haizhou Li and Tomi Kinnunen and Khe Chai Sim, "Using discrete probabilities with Bhattacharyya measure for SVM- based speaker verification", IEEE Transaction on audio, speech and language processing, vol. 19, no. 4, May 2011. cs.uef.fi/sipu/pub/Bhattacharyya_TASL.pdf
[9] Pedro.J, Moreno, Purdy.P, Ho, "Verification Using Probabilistic Distance Kernels", Eurospeech, 1-4 September, 2003. www.hpl.hp.com/techreports/2004/HPL-2004-7.html
[10] Man Wai Mak, Wei Rao, "Utterance Partitioning with acoustic vector resampling for GMM-SVM speaker verification", Speech Communication, 53, pp. 119-130, 2011. http://citeseerx.ist.psu.edu/viewdoc/citations;jsessionid=524637905DED778C54349B0B2BB0FEC7?doi=10.1.1.170.4398
[11] Zhao Jian, Dong Yuan, Zhao Xianyu, "Score Normalization and Language Robustness in Text Independent Multi-language Speaker Verification", Lecture notes in computer science, 2007, pp.4681:1121-1130. ieeexplore.ieee.org/iel5/5971803/6074142/06074159.pdf
[12] Yao, Y., Frasconi, P., & Pontil, M. (2001). Fingerprint classification with combinations of support vector machines, AVBPA 2001, LNCS 2091, pp. 253–258. link.springer.com/chapter/10.1007/3-540-45344-X_37
[13] Campbell, W. M., Singer, E., Torres-Carrasquillo, P. A., & Reynolds, D. A. (2004). Language recognition with support vector machines, Odyssey 2004, May 31st– June 4th, Toledo, Spain. https://www.ll.mit.edu/mission/cybersec/...files/.../040531_CampbellW_SingerE.pdf
[14] S.Zribi Boujelbene, D.Ben Ayed Mezghani, and N. Ellouze, "Improved Feature data for Robust Speaker Identification using hybrid Gaussian Mixture Models - Sequential Minimal Optimization System", The International Review on Computers and Software, Vol. 4.3, ISSN: 1828-6003, May 2009, pp.344-350.
[15] W. Zunjingand, and C. Zhigang, "Improved MFCC-Based Feature for Robust Speaker Identification", Tsinghua Science and Technology, Vol. 10.2, 2005, pp. 158-161. http://whale.cse.nsysu.edu.tw/~khwu/Improved%20MFCC-Based%20Feature%20for%20Robust%20Speaker%20Identification.pdf
[16] Bing Xiang, Toby Berger, "Efficient Text-Independent Speaker Verification With Structural Gaussian Mixture Models And Neural Network", IEEE Transactions On Speech And Audio Processing, Vol. 11, No. 5, pp. 447-456, September 2003.http://www.dl.edi-info.ir/Efficient%20Text-Independent%20Speaker%20Verification%20with%20Structural%20Gaussian%20Mixture%20Models.pdf
[17] D.A.Reynolds, T.F.Quatieri, R.B.Dunn, "Speaker verification using adapted Gaussian mixture models", Digital Signal Process. 10, 19–41, 2000. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.556.4188&rep=rep1&type=pdf
[18] Fuqian Shi, Jiang Xu, "Emotional Cellular-Based Multi- Class Fuzzy Support Vector Machines on Product's KANSEI Extraction", International Journal of Applied Mathematics & Information Sciences, Vol.6, No. 1, pp. 41-49, 2012. https://pdfs.semanticscholar.org/a184/1cd6b389bb36340524a63481a4ddab564c84.pdf
[19] C.F.Lin and S.D.Wang, "Fuzzy support vector machines," IEEE Transactions on Neural Networks, vol. 13, no. 2, March 2002.
[20] Chun-fu Lin, Sheng-de Wang, "Training algorithms for fuzzy support vector machines with noisy data", Pattern Recognition Letters, 25, pp. 1647–1656, 2004. www.cs.pdx.edu/~mm/aml2010/JoshHoakPresentation.pdf

[21] C.J.C.Burges and B.Schölkopf, "Improving the Accuracy and Speed of Support Vector Learning Machines", Advances in Neural Information Processing Systems 9, Cambridge, MIT Press, 1997, pp. 375–381. https://people.eecs.berkeley.edu/~malik/cs294/decoste-scholkopf.pdf
[22] W.M.Campbell, J.P.Campbell, D.A.Reynolds, D.A.Jones and T.R.Leek, "High Level Speaker Verification with Support Vector Machines", in Proceedings of the International Conference on Acoustics Speech and Signal Processing,2004, pp.73-76.
[23] S.Furui, "Cepstral analysis technique for automatic speaker verification", IEEE Trans. Acoust. Speech Signal Processing, vol.ASSP-29, pp.254-272, Apr.1981. https://www.researchgate.net/publication/3176892_Cepstral_analysis_technique_for_automatic_speaker_verification
[24] H.Hermansky and N.Morgan, "RASTA processing of speech", IEEE Trans. Speech and Audio Processing, vol.2, no.4, pp.578-589, 1994. ieeexplore.ieee.org/document/326616/
[25] S.Chen and R.A.Gopinath, "Gaussianization", Proc. NIPS 2000, Denver Colorado.

P.Shanmugapriya is an associate professor in the department of ECE, Saranathan College of Engineering, Tamil Nadu, India. She obtained M.Tech from NIT, Trichy, in the discipline of Master of Communication systems in the year of 2005. She has completed her PhD from the faculty of Information and Communication, Anna University, Chennai, Tamil Nadu, India in 2015. Her fields of interests include Speech Processing, Soft computing and Pattern recognition.

Dr.Y.Venkataramani is a Professor and Dean (R&D), Saranathan College of Engineering, Tamil Nadu, India. He has got more than 40 years of experience in the field of teaching. He has guided several PhDs in the area of Electronics and Communication engineering. His fields of interests include Computer Networking, soft computing and Pattern classification.

DOI : 10.24297/jac.v13i10.5839
www.cirworld.com