Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009

Design of a Modified One-Against-All SVM Classifier

J.Manikandan and B.Venkataramani
Department of ECE, National Institute of Technology, Trichy, (NITT), India.
{jmanikandan.nitt@gmail.com, bvenki@nitt.edu}

***Abstract*—Support Vector Machine (SVM) is one of the state of-the-art tools for linear and nonlinear pattern classification. One of the design issues in SVM classifier is reducing the number of support vectors without compromising the classification accuracy. In this paper, a novel technique which requires only a subset of the support vectors is proposed. The subset is obtained by including only those support vectors for which Lagrange multiplier is greater than a threshold. In order to find the subset which yields the highest classification accuracy with the least number of support vectors in the subset, the recognition performance corresponding to subsets with different threshold values are to be evaluated and compared. The proposed technique is applied for SVM based isolated digit recognition system and is studied using speaker dependent as well as multispeaker dependent TI46 database of isolated digits. Two feature extraction techniques, one using LPC and another using MFCC are applied to the speech from the above database and the features are mapped using SOFM. This in turn is used by the SVM classifier to evaluate the recognition accuracy. The proposed technique is applied to One-Against-All (OAA) scheme and is denoted as Modified One-Against-All (M-OAA) approach in this paper. Based on this study, it is found that for MFCC feature input, the proposed M-OAA based SVM classifier approach results in reduction of support vectors by a factor of 1.86 to 18.3 with no compromise in recognition accuracy. For LPC feature input, the M-OAA based SVM classifier results in reduction of support vectors by a factor of 1.59 to 2.52 without any compromise in recognition accuracy for some cases and with a maximum of 1% degradation in recognition accuracy for some cases. The proposed approach is also applicable for other schemes such as Half-Against-Half (HAH) and Directed Acyclic Graphs (DAG) based SVM classifiers as well as for any other classification problem such as face recognition, fingerprint recognition, target recognition, speaker recognition and speaker verification.**

***Keywords*—SVM, Support Vectors, OAA, Machine Learning, Pattern Recognition, Isolated Digit Recognition**

I. INTRODUCTION

Support Vector Machine (SVM) is one of the popular techniques for pattern recognition and is considered to be the state-of-the-art tool for linear and nonlinear classification [1]. SVM is basically a binary classifier and it has been employed for several applications such as beam forming [2], ultra wide band (UWB) channel equalization [3], channel estimation in Orthogonal Frequency Division Multiplexing (OFDM) systems [4], voice activity detection [5] and target recognition [6]. The SVM classifier has been proposed for binary classification in literature and it has been extended for the design of multiclass SVM classifiers [7].

Computational cost and the classification time for SVM based pattern recognition systems depend on the number of support vectors required for the design of SVM classifier and the kernel employed. Increase in number of support vectors lead to increase in computational requirements such as floating point multiplication and addition, which makes the classification slow and hence, it is crucial to decrease the number of support vectors. K-means clustering technique is proposed in [8] for reducing the number of support vectors of the SVM classifier for handwritten digits. In order to reduce the training time for SVM classifier, number of techniques such as chunking[9], decomposition algorithm [10], Sequential Minimal Optimization (SMO) technique [11] and online support vector classifier [12] are reported in the literature.

In this paper, a novel technique, which enables a subset of the support vectors to be used with SVM classifier, is proposed in order to reduce the SVM classification time, computational cost and also the memory required for storing the parameters. The subset is obtained by including those support vectors for which Lagrange multiplier is greater than a threshold and the efficacy of this technique is evaluated by conducting various experiments using an SVM based isolated digit recognition system.

The organization of the paper is as follows: Section II gives an overview of the algorithm used in Support Vector Machine (SVM) and the architecture of SVM based isolated digit recognition system. In Section III, the proposed M-OAA based SVM classifier is explained. The performance results of proposed approach are reported in Section IV, followed by conclusion and references.

II. SVM CLASSIFIER FOR ISOLATED DIGIT RECOGNITION SYSTEM

A. SVM Classifier

The basic form of a SVM classifier can be expressed as:

$$Y(z) = \boldsymbol{w} \cdot \boldsymbol{\Phi}(z) + b \qquad (1)$$

978-1-4244-2794-9/09/$25.00 ©2009 IEEE

where $\boldsymbol{z}$ is the test input vector, $\boldsymbol{w}$ is a vector normal to the hyper-plane which separates the classes in the feature space. The feature space is produced from the feature mapping function $\boldsymbol{\Phi}(\cdot)$ which can be either linear or non-linear. b is the bias. The separating hyper-plane (described by $\boldsymbol{w}$) is determined by minimizing the structural risk instead of the empirical error [13]. Minimizing the structural risk is equivalent to seeking the optimal margin between two classes and the width of the margin is $\frac{2}{\boldsymbol{w}.\boldsymbol{w}}$.

Let $\{\boldsymbol{x}_i, \boldsymbol{y}_i\}$ for i=1,2,…N denote the training data set where $\boldsymbol{y}_i$ is the target output for training data $\boldsymbol{x}_i$. SVM training can be posed as the constrained optimization problem which maximizes the width of the margin and minimizes the structural risk and is given by

$$\min_{\boldsymbol{w},b} \ \frac{1}{2}\boldsymbol{w}.\boldsymbol{w} + \mathrm{C}\sum_{i=1}^{N}\xi_i \qquad (2)$$

subject to

$$\boldsymbol{y}_i(\boldsymbol{w} \,.\, \boldsymbol{\Phi}(\boldsymbol{x}_i) + b) \geq 1 - \xi_i$$
$$\xi_i \geq 0, \ \forall\, i$$

where C is the trade-off parameter and ξ_i is the slack variable, which measures the deviation of a data point from the ideal condition of pattern separability. The penalty parameter "C" controls the trade-off between the complexity of the decision function and the number of wrongly classified testing points. The correct C value cannot be known in advance and a wrong choice of the SVM penalty parameter C can lead to a severe loss in performance. Therefore, the parameter values are usually estimated from the training data by cross-validation and exponentially growing sequences of C.

The solution to (2) can be reduced to a QP optimization problem [17]:

$$\max_{\boldsymbol{\alpha}} \ \frac{1}{2}\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{\alpha} - \frac{1}{2}\boldsymbol{\alpha}^{\mathrm{T}}\mathrm{H}\,\boldsymbol{\alpha} \qquad (3)$$

subject to

$$0 \leq \alpha_i \leq \mathrm{C}, \ \forall\, i,$$
$$\sum_{i=1}^{N}\alpha_i y_i = 0,$$

where $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots\ldots \alpha_N]^{\mathrm{T}}$, and H is a NxN matrix, called the kernel matrix with $(i,j)^{th}$ element given by

$$\mathrm{H}(i,j) = \boldsymbol{y}_i \boldsymbol{y}_j\, \boldsymbol{\Phi}(\boldsymbol{x}_i).\, \boldsymbol{\Phi}(\boldsymbol{x}_j) \qquad (4)$$

There is a Lagrange multiplier α_i for each training sample $\boldsymbol{x}_i$. Those samples whose α_i's are nonzero are called Support Vectors (SV) and a portion of training samples become SVs[16]. Solving the QP problem [18] yields:

$$\boldsymbol{w} = \sum_{i=1}^{N}\alpha_i y_i\ \boldsymbol{\Phi}(\boldsymbol{x}_i) \qquad (5)$$

and (5) can be rewritten in terms of support vectors as :

$$\boldsymbol{w} = \sum_{sv=1}^{NSV}\alpha_{sv} y_{sv}\ \boldsymbol{\Phi}(\boldsymbol{x}_{sv}), \qquad (6)$$

where NSV denotes the number of support vectors and the parameters α_{sv}, y_{sv} and $\boldsymbol{x}_{sv}$ represent the parameters corresponding to each support vector. From (1) and (6), the SVM classifier equation for a test data $\boldsymbol{z}$ is expressed as:

$$\begin{aligned} Y(\boldsymbol{z}) &= \sum_{sv=1}^{NSV}\alpha_{sv} y_{sv}\ \boldsymbol{\Phi}(\boldsymbol{x}_{sv})\,.\boldsymbol{\Phi}(\boldsymbol{z}) + b \\ &= \sum_{sv=1}^{NSV}\alpha_{sv} y_{sv}\ \mathrm{K}(\boldsymbol{x}_{sv}\,.\,\boldsymbol{z}) + b, \end{aligned} \qquad (7)$$

where $\mathrm{K}(\boldsymbol{x}_{sv}.\boldsymbol{z}) = \boldsymbol{\Phi}(\boldsymbol{x}_{sv}).\boldsymbol{\Phi}(\boldsymbol{z})$ is a kernel function The advantage of using kernel function is that the $\boldsymbol{\Phi}(.)$ need not be found explicitly for each kernel. It is observed from (7) that the increase in number of support vectors lead to increase in computational requirements and hence making the classifier slow for non-linear kernels. Examples of some of the most commonly-used kernel functions [14][16] in various SVM classifiers are given in Table I.

TABLE I. EXAMPLES OF SVM KERNELS WITH THEIR KERNEL FUNCTION

Linear	$\mathrm{K}(\boldsymbol{x}\,.\,\boldsymbol{z}) = \boldsymbol{x}_i\,.\,z_i$
Polynomial	$\mathrm{K}(\boldsymbol{x}\,.\,\boldsymbol{z}) = (\boldsymbol{x}_i\,.\,z_i + \gamma)^d$, where d is the degree of polynomial
RBF	$\mathrm{K}(\boldsymbol{x}\,.\,\boldsymbol{z}) = \exp(-\gamma\lVert\boldsymbol{x}_i - z_i\rVert^2)$
Sigmoid	$\mathrm{K}(\boldsymbol{x}\,.\,\boldsymbol{z}) = \tanh(\gamma\, \boldsymbol{x}_i\,.z_i + 1)$
Wavelet	$\mathrm{K}(\boldsymbol{x}\,.\,\boldsymbol{z}) = \prod_{i=1}^{N}\mathrm{h}(\frac{x_i - z_i}{a})$, where $\mathrm{h}(x) = \cos(1.75x)\exp(-\frac{x^2}{2})$ and a is the dilation parameter

For a multi-class classification problem, the output domain is $Y(\boldsymbol{z})$ = {1,2,….m} for a m-class problem. Support vectors, (α_i) and bias (b_i) values are associated to each of the m-classes, where i ε{1,2,....m}, and the decision function is given by

$$Y(\boldsymbol{z}) = \mathrm{argmax}_i(Y_i(\boldsymbol{z})),\ \forall\, i, \qquad (8)$$

where $Y_i(\boldsymbol{z})$ is computed for each class i using (7). Geometrically this is equivalent to associating a hyperplane to each class, and assigning the test input $\boldsymbol{z}$ to that class whose hyperplane is furthest from it. More details about SVM classifier can be had from [13], [14] and [17].

B. *Isolated Digit Recognition*

Isolated digit recognition is a problem of pattern classification which has a very vast feature size. Fig.1 shows the block diagram of proposed isolated digit recognition system using SVM classifier, comprising feature extraction, self organized mapping of the features extracted and SVM classifier architecture stage. The input speech is divided into frames and LPC/MFCC coefficients are computed for each frame. (The Linear Predictive Coefficient (LPC) features are extracted from the input speech using the pre-emphasis, frame blocking, windowing, and autocorrelation blocks. Details about the steps involved in extracting the LPC features can be had from [18],[19]. Similarly, Mel Frequency Cepstral Coefficients (MFCC) are computed using the following blocks: pre-emphasis, frame blocking, windowing, DFT, Mel filter bank processing, IDFT, computation of energy, computation of

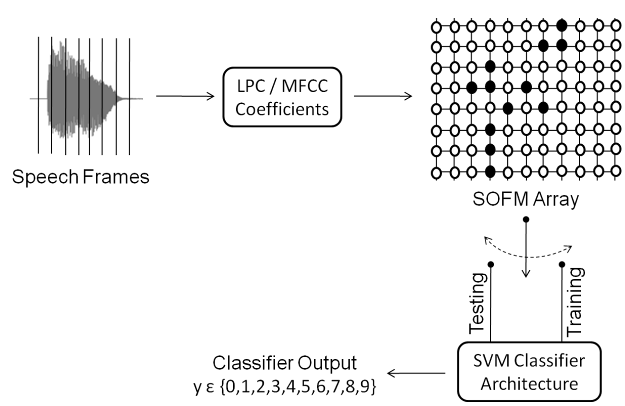

Figure 1. Block diagram of SVM based Isolated Digit Recognition

delta and double delta features. Detailed account of the procedure adopted for computation of MFCC is given in [20].) Each frame results in 11 LPC coefficients or 25MFCC (13MFCC + 12 Delta MFCC) coefficients. The total number of LPC/MFCC coefficients for each digit is not uniform because each input digit comprises of different number of frames. SVM based recognition requires equal number of feature vectors for each input. In order to make the size of input feature uniform for all the digits, Kohonen's self-organizing feature map (SOFM) [21] is used. SOFM reduces the input feature size for classification too. The SOFM transforms the acoustic vector sequences of features into trajectories in a square matrix of fixed dimension, where each node of the matrix take on binary values [22]. The feature map is trained by Kohonen's self-organization learning. For the purpose of SOFM training, ten utterances of all the ten digits are used. The self-organization learning has the property that after training, physically similar input vectors in Euclidean space, correspond to topologically close nodes in the feature map. The size of SOFM can be 12x12, 18x18, 24x24 or any other combination. Mapped features of SOFM for each digit are obtained using the SOFM weights. In the SOFM array shown in Fig.1, black dots, white dots denote the values '1' and '0' respectively.

III. Motivation for M-OAA based SVM Classifier

Support vectors are the data points lying on the margins of the decision boundary that are employed in forming the decision boundary for the SVM based multi-class pattern recognition systems. As noted in Section II.A, those training samples whose Lagrange multpliers (α_i's) are nonzero are called Support Vectors (SV) and a portion of training samples become SVs. It may be noted from (7) that the computational cost of the SVM classifier depends on the number of support vectors (NSV). Support Vectors are obtained during the learning phase of SVM based recognition system. One of the popular and simplest techniques used for multi-class classifier is the one-against-all (OAA) algorithm which is also called as one-against-rest or one-against-remaining. The decision boundary adapted for the OAA-based SVM classifier for isolated digit recognition system is shown in Fig.2. In Fig.2, the entire training dataset is used for forming a decision boundary between classes and the support vectors are obtained for each class. A parallel SVM classifier may be designed as shown in Fig.3. The advantage of this approach is that the boundary is always stringent with respect to each class and hence each classifier may recognize the input test data more accurately. The disadvantage of this technique is that the number of support vectors may be increased due to the stringent boundary between classes.

A study of the OAA based SVM classifier for isolated digit recognition system is carried out using both the speaker dependent and multispeaker-dependent TI46 database [23]. For speaker dependent case, ten utterances of each digit from 1 female speaker are used for training the system and ten utterances of test data for each digit from the same speaker at different sessions are used for testing. For multispeaker-dependent case, the system is trained using ten utterances of each digit

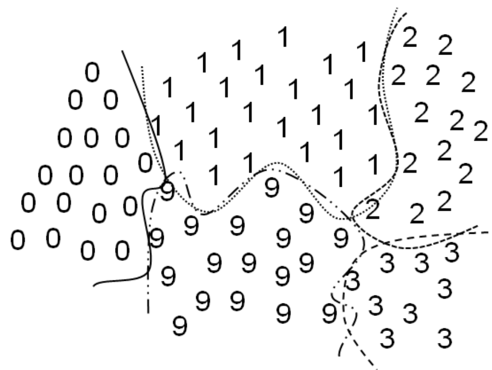

Figure 2. Learning Mechanism for SVM-OAA based recognition system

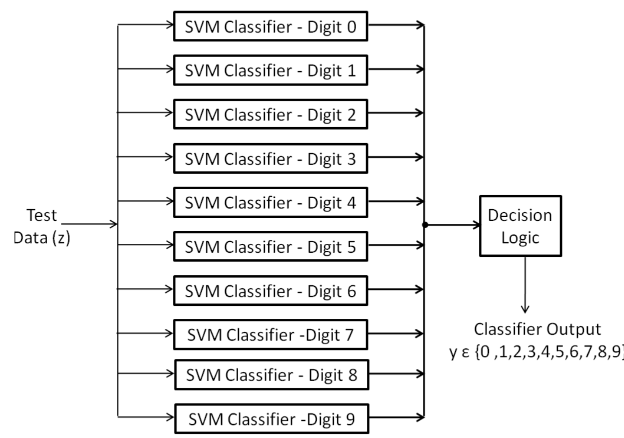

Figure 3. Parallel SVM Classifier

from 3 speakers, and is tested using the testing tokens of each speaker, respectively. Feature extraction using LPC/MFCC and SOFM feature mapping have been implemented using C-code. The SOFM output data is fed as input to the SVM classifier during training and classification phase of the system. The SVM classification algorithm described in Section II is implemented using MATLAB [24] as well as C-code and the performance of the SVM based isolated digit recognition system is evaluated with LPC and MFCC feature inputs using various kernels.

The recognition accuracy (RA) achieved and the number of support vectors (SVs) required for the SVM based isolated digit recognition system for various SOFM sizes with LPC feature input and speaker dependent case are tabulated in Table II. It is observed that for both linear and non-linear kernels, the best recognition accuracy is achieved with SOFM of size 18x18 which coincides with the result in [22] and hence SOFM of size 18x18 is used for further analysis.

It is observed that linear kernel and wavelet kernel requires 692 and 800 support vectors respectively to form the decision boundary. This also implies that there are around 692 Lagrange multipliers (with linear kernel, C=4) and 800 Lagrange multipliers (with wavelet kernel, C=2) having $\alpha > 0$. During in-depth analysis, it is observed that there are several Lagrange multipliers, which are nearer to zero and are relatively small when compared with other Lagrange multipliers. The values of Lagrange multipliers for all the support vectors obtained using linear kernel and wavelet kernel are tabulated in Table III. It may be observed that for linear kernel, 42.2% of support vectors fall in the range $0 < \alpha \leq 0.00001$ and 51.88% of support vectors fall in the range $0 < \alpha \leq 0.01$. It may be observed that for wavelet kernel, 25.87% of support vectors fall in the range $0 < \alpha \leq 0.00001$ and 63.12% of support vectors fall in the range $\alpha > 0.1$. In order to assess the importance of those support vectors whose Lagrange multipliers are relatively small, the performance of the SVM classifier which uses only a subset of the support vectors corresponding to those Lagrange multipliers whose value is greater than a threshold (denoted as $\alpha_{THRESHOLD}$) is evaluated for wavelet kernel with MFCC feature input and the result is given in Table IV. From Table IV, it is found that it is possible to reduce the number of support vectors without compromising the recognition accuracy. Based on this, the novel technique for SVM classifier which uses only a subset of support vectors is proposed.

TABLE II. RECOGNITION PERFORMANCE FOR VARIOUS SOFM SIZES

Kernel	SOFM Size	Recognition Accuracy (RA)	No. of Support Vectors (SVs)
Linear	12x12	85%	627
	18x18	**97%**	**692**
	24x24	94%	739
Non-Linear Wavelet(a=10)	12x12	89%	703
	18x18	**99%**	**800**
	24x24	92%	865

TABLE III. BREAK-UP OF SUPPORT VECTORS INTO DIFFERENT RANGES OF LAGRANGE MULTIPLIER VALUE WITH LPC FEATURE INPUT

Range of Lagrange Multiplier values ($\alpha > 0$)	Linear Kernel SVs #	Wavelet Kernel SVs #
(0,0.00001]	292	207
(0.00001,0.01]	67	13
(0.01,0.1]	270	75
(0.1,1]	63	381
(1,C]	0	124
Total	**692**	**800**

TABLE IV. PERFORMANCE OF SVM CLASSIFIER WITH DIFFERENT VALUES OF $\alpha_{THRESHOLD}$ FOR WAVELET KERNEL WITH MFCC FEATURE INPUT

$\alpha_{THRESHOLD}$	C	Recognition Accuracy (RA)	No. of Support Vectors (SVs#)
0	1	100	785
0.1	1	100	384
0.2	1	100	268
0.3	1	100	173
0.4	1	100	142
0.5	1	100	117
0.6	1	100	96
0.7	**1**	**100**	**85**
0.8	1	99	80

IV. PERFORMANCE ANALYSIS OF M-OAA BASED SVM CLASSIFIER

One-Against-All (OAA) SVM classifier using a subset of support vectors is denoted as modified One-Against-All (M-OAA) SVM classifier. The performance of the M-OAA SVM classifier is evaluated for different values of '$\alpha_{threshold}$'. The recognition performance is obtained for various kernels using LPC and MFCC feature inputs. The performance results corresponding to speaker dependent case and multi speaker dependent case are given in Table V and Table VI respectively. Factor by which Support Vectors are reduced on using M-OAA based SVM Classifier is given in Table VII. The following observations may be made from Table V, Table VI and Table VII.

- The recognition performance of OAA based SVM classifier corresponds to the results in the first rows of Table V and Table VI. These rows correspond to $\alpha > 0$ and these rows yield the support vectors for various kernels as per the definition in the literature.
- The optimum threshold value is defined as the maximum value of $\alpha_{THRESHOLD}$ for which SVM classifier requires the least number of support vectors with either no compromise or negligible compromise on recognition accuracy.
- The results corresponding to the optimal threshold values are indicated in bold-face.
- For MFCC, the optimum threshold value ($\alpha_{THRESHOLD}$) results in lesser number of support vectors without any degradation in recognition accuracy for both speaker dependent and multispeaker dependent cases.
- For LPC, the optimum threshold value ($\alpha_{THRESHOLD}$) results in lesser number of support vectors without any degradation in recognition accuracy for speaker dependent case (on using linear and polynomial kernels) and multispeaker dependent case (on using RBF and wavelet kernels). For the remaining kernels in both cases, a maximum of 1% degradation in recognition accuracy is observed with a reduction of support vectors by a factor of 1.59 to 2.52.
- It may be noted that the value of optimum threshold lies in the range (0, C] and it varies for each kernel and input data type.
- Significant reduction in number of support vectors is observed for all the cases by choosing an optimum threshold value. It may be observed that the M-OAA based SVM classifier reduces the number of support vectors by a factor ranging from 1.59 to 18.3. The sigmoid kernel could not classify for the multispeaker dependent test case and hence it is left blank.

V. CONCLUSION

In this paper, a novel technique is proposed which enables a subset of the support vectors to be used with SVM classifier in order to reduce the SVM classification time and the memory required for storing the parameters. The above approach is validated using SVM-OAA based isolated digit recognition system for both speaker-dependent and multispeaker-dependent cases. The number of support vectors required for classification is significantly reduced without compromising

TABLE V. SPEAKER DEPENDENT TEST : RECOGNITION PERFORMANCE OF MODIFIED OAA BASED SVM CLASSIFIER USING VARIOUS KERNELS

$\alpha_{THRESHOLD}$ for LPC feature input	Linear			Poly (deg=2)			RBF (γ=1/324)			Sigmoid (γ=1/324)			Wavelet (a=10)		
	C	RA	SVs#	C	RA	SVs#	C	RA	SVs#	C	RA	SVs#	C	RA	SVs#
0	5	97	692	1	92	881	10	98	715	10	98	674	2	99	800
0.00001	1	97	400	1	92	729	10	97	432	100	97	398	10	98	593
0.0001	1	97	396	**1**	**92**	**387**	10	97	432	100	97	398	10	98	593
0.001	**1**	**97**	**374**	100	91	729	10	97	432	100	97	398	10	98	593
0.01	10	95	374	1000	91	729	10	97	432	100	97	398	10	98	590
0.1	100	95	374	-	-	-	10	97	423	100	97	398	10	98	574
1	-	-	-	-	-	-	10	97	358	100	97	391	**10**	**98**	**424**
2	-	-	-	-	-	-	**10**	**97**	**284**	100	97	382	10	97	306
3	-	-	-	-	-	-	10	96	231	100	97	380	-	-	-
4	-	-	-	-	-	-	10	96	197	100	97	372	-	-	-
5	-	-	-	-	-	-	-	-	-	100	97	365	-	-	-
6	-	-	-	-	-	-	-	-	-	100	97	358	-	-	-
7	-	-	-	-	-	-	-	-	-	100	**97**	**353**	-	-	-
8	-	-	-	-	-	-	-	-	-	100	96	342	-	-	-
$\alpha_{THRESHOLD}$ for MFCC feature input	**Linear**			**Poly (deg=2)**			**RBF (γ=1/324)**			**Sigmoid (γ=1/324)**			**Wavelet (a=10)**		
	C	**RA**	**SVs#**	**C**	**RA**	**SVs#**	**C**	**RA**	**SVs#**	**C**	**RA**	**SVs#**	**C**	**RA**	**SVs#**
0	1	100	691	1	99	915	1	100	740	1	100	700	1	100	785
0.00001	1	100	378	1	99	790	1	100	423	10	100	391	1	100	572
0.0001	1	100	373	1	99	467	1	100	423	10	100	391	1	100	572
0.001	1	100	343	**1**	**99**	**50**	1	100	422	10	100	391	1	100	570
0.01	**1**	**100**	**149**	500	99	737	1	100	414	10	100	390	1	100	556
0.1	10	100	149	5000	99	737	1	100	350	10	100	381	1	100	384
0.2	-	-	-	-	-	-	**1**	**100**	**289**	**10**	**100**	**376**	1	100	268
0.3	-	-	-	-	-	-	1	98	287	10	99	371	1	100	173
0.4	-	-	-	-	-	-	-	-	-	-	-	-	1	100	142
0.5	-	-	-	-	-	-	-	-	-	-	-	-	1	100	117
0.6	-	-	-	-	-	-	-	-	-	-	-	-	1	100	96
0.7	-	-	-	-	-	-	-	-	-	-	-	-	1	**100**	**85**
0.8	-	-	-	-	-	-	-	-	-	-	-	-	1	99	80

TABLE VI. MULTI-SPEAKER DEPENDENT TEST : RECOGNITION PERFORMANCE OF MODIFIED OAA BASED SVM CLASSIFIER USING VARIOUS NON-LINEAR KERNELS

$\alpha_{THRESHOLD}$ for LPC feature input	Poly (deg=2)			RBF (γ=1/50)			Wavelet (a=7)		
	C	RA	SVs#	C	RA	SVs#	C	RA	SVs#
0	3	92.67	2340	5	92.33	1858	1	94.67	2342
0.00001	1	92.33	1667	15	92.67	1005	1	94.67	1681
0.0001	**2**	**92.33**	**1471**	15	92.67	1005	1	94.67	1680
0.001	50	91.33	1604	15	92.67	1005	1	94.67	1679
0.01	500	91.33	1604	15	92.67	1005	1	94.67	1642
0.1	5000	91.33	1604	15	92.67	998	2	94.67	1435
0.2	-	-	-	15	92.67	989	3	94.67	1357
0.3	-	-	-	15	92.67	982	4	94.67	1309
0.4	-	-	-	**15**	**92.67**	**969**	5	94.67	1287
0.5	-	-	-	15	92.33	960	**6**	**94.67**	**1272**
0.6	-	-	-	-	-	-	8	94.67	1309
$\alpha_{THRESHOLD}$ for MFCC feature input	**Poly (deg=2)**			**RBF (γ=1/20)**			**Wavelet (a=7)**		
	C	**RA**	**SVs#**	**C**	**RA**	**SVs#**	**C**	**RA**	**SVs#**
0	1	100.0	2306	1	100.0	2251	1	100.0	2153
0.00001	1	100.0	1511	1	100.0	1466	1	100.0	1303
0.0001	1	100.0	1102	1	100.0	1464	1	100.0	1302
0.001	**10**	**100.0**	**1102**	1	100.0	1458	1	100.0	1299
0.01	100	100.0	1102	1	100.0	1410	1	100.0	1269
0.1	1000	100.0	1102	**1**	**100.0**	**921**	**1**	**100.0**	**870**
0.2	2000	100.0	1102	2	100.0	921	2	100.0	870

TABLE VII. FACTOR BY WHICH SUPPORT VECTORS ARE REDUCED ON USING M-OAA BASED SVM CLASSIFIER

Dataset →	Spk. Dep.		Multi Spk. Dep.	
Kernel ↓	**MFCC**	**LPC**	**MFCC**	**LPC**
Polynomial	18.3	2.28	2.09	1.59
Wavelet	9.24	1.89	2.47	1.84
RBF	2.56	2.52	2.44	1.92
Sigmoid	1.86	1.91	-	-

the recognition accuracy for MFCC feature inputs. The proposed approach can also be employed for any other SVM classifier architecture such as HAH and DAG as well as for any other classification problem such as face recognition, fingerprint recognition, target recognition, speaker recognition and speaker verification.

REFERENCES

[1] Christopher J.C. Burges, "A tutorial on Support Vector Machines for Pattern Recognition," Data Mining and Knowledge Discovery 2, 121-167,1998.

[2] M. Martínez Ramón, Nan Xu, and C. G. Christodoulou, "Beamforming using Support Vector Machines," IEEE Antennas And Wireless Propagation Letters, Vol. 4, 2005, pp. 439-442.

[3] Mohamed S. Musbah and Xu Zhu, "Support Vector Machines for DS-UWB Channel Equalisation," Department of Electrical Engineering & Electronics, University of Liverpool, UK, IEEE,2007, pp. 524-527.

[4] M. Julia Fernández-Getino García, José Luis Rojo-Álvarez, "Support Vector Machines for Robust Channel Estimation in OFDM," IEEE Signal Processing Letters, Vol. 13, No. 7, July 2006, pp. 397-400.

[5] Fengyan Q, Changchun Bao and Yan Liu, "A Novel Two-Step SVM Classifier for voiced/unvoiced/silence classification of speech," in Proc. ISCSLP 2004, pp. 77-80.

[6] Qun Zhao and Jose C. Principe, "Support Vector Machines for SAR Automatic Target Recognition," IEEE Trans. Aerosp. Electron. Syst., Vol. 37, No. 2 April 2001, pp.643-654.

[7] Andrew W. Moore, "Support Vector Machines," School of Computer Science,Carnegie Mellon University.

[8] Jiaqi Wang, Xindong Wu and Chengqi Zhang, "Support vector machines based on K-means clustering for real-time business intelligence systems," Int. J. Business Intelligence and Data Mining, Vol. 1, No. 1, 2005, pp. 54-64.

[9] Joachims, T., "Making Large-scale SVM Learning Practical," Advances in Kernel Methods - Support Vector Learning, MIT Press, Cambridge, MA, 1999, pp. 169–184.

[10] Osuna E., Freund R. and Girosi F., "Training Support Vector Machines," in Proc. Conf. Computer Vision and Pattern Recognition, CVPR, 1997, pp. 130–136.

[11] John C. Platt, "Sequential Minimal Optimization: A Fast Algorithm for training support vector machines," Technical Report MSR-TR-98-14, Microsoft Research, 1998.

[12] K.W. Lau and Q.H. Wu, "Online training of support vector classifier," Pattern Recognition 36, 2003, pp.1913-1920.

[13] Nello Cristianini and John Shawe-Taylor, An introduction to Support Vector Machines and other kernel-based learning methods, Cambridge University Press, 2000

[14] Simon Haykin, Neural Networks, Prentice Hall of India, Second Edition, 2003.

[15] Edwin K.P.Chong and Stanislaw H. Zak, An introduction to Optimization, Wiley Interscience Publications, 2004

[16] Li Zhang, Weida Zhou, and Licheng Jiao, "Wavelet Support Vector Machine," IEEE Trans. Systems, Man, and Cybernetics—Part B: Cybernetics, Vol. 34, No. 1, February 2004.

[17] Corrina Cortes and V Vapnik, "Support Vector Networks", J. of Machine Learning, 1995.

[18] Ben Gold amd Nelson Morgan, Speech and Audio Signal Processing: Processing and Perception of Speech and Music, John Wiley and Sons Inc, 2000.

[19] Lawrence Rabiner and Biing-Hwang Juang, Fundamentals of Speech Recognition, Prentice Hall Signal Processing Series, 1993.

[20] Daniel Jurafsky and James H. Martin, Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, Prentice-Hall New Jersey, 2008.

[21] Martin T Hagan, Howard B. Demuth and Mark Beale, Neural Network Design, PWS Publishing Company, 1996.

[22] Zezhen Huang and Anthony Kuh, "A Combined Self-Organizing Feature Map and Multilayer Perceptron for Isolated Word Recognition," IEEE Trans. Signal Proc., Vol.40, No. 11, pp.2651-2657, Nov 1992.

[23] http://www.ldc.upenn.edu/Catalog/readme_files/ti46.readme.html

[24] The Mathworks Inc. – MATLAB, www.mathworks.com.