



## Improved Admissible Kernel Function for Support Vector Machines in Banach Space for Multiclass Data

R.Sangeetha, B.Kalpana

Ph.D Research Scholar, Avinashilingam University,Coimbatore - 43.

sangeethadj@gmail.com

Professor,Avinashilingam University,Coimbatore - 43.

kalpanabsekar@yahoo.com

### ABSTRACT

Classification based on supervised learning theory is one of the most significant tasks frequently accomplished by so-called Intelligent Systems. Contrary to the traditional classification techniques that are used to validate or contradict a predefined hypothesis, kernel based classifiers offer the possibility to frame new hypotheses using statistical learning theory (Sangeetha and Kalpana, 2010). Support Vector Machine (SVM) is a standard kernel based learning algorithm where it improves the learning ability through experience. It is highly accurate, robust and optimal kernel based classification technique that is well-suited to many real time applications. In this paper, kernel functions related to Hilbert space and Banach Space are explained. Here, the experimental results are carried out using benchmark multiclass datasets which are taken from UCI Machine Learning Repository and their performance are compared using various metrics like support vector, support vector percentage, training time and accuracy.

### Keywords

Multiclass Classification,Admissible Kernel Function,Support Vector Machine, Banach Space,Lévy Distribution.

### Academic Discipline And Sub-Disciplines

Computer Science

### SUBJECT CLASSIFICATION

Computer Science Subject Classification

### TYPE (METHOD/APPROACH)

Experimental Analysis

# Council for Innovative Research

Peer Review Research Publishing System

**Journal:** INTERNATIONAL JOURNAL OF COMPUTERS & TECHNOLOGY

Vol 11, No.2

[editor@cirworld.com](mailto:editor@cirworld.com)

[www.cirworld.com](http://www.cirworld.com), [member.cirworld.com](http://member.cirworld.com)

## 1. INTRODUCTION

Despite of about 50 years of research investigation, developing a unified theory for a pattern recognizer has been an elusive task (Dutt et al. 2011). Three processes in pattern recognition are data acquisition, data analysis and pattern classification. Data acquisition is a first step in which the acquired raw data is converted into information, where it can be given as an input to the computing machine for next step in processing. Data analysis is a process where the collected information and pattern classes are analyzed. Based on the knowledge derived from data analysis step, pattern classification categorizes the data into different classes.

Data that are given as an input to the pattern classification is splitted into two types of datasets i.e. training dataset and testing dataset. Pattern classification (Sharma and Kishor, 2007) technique first takes the training dataset as an input to train the system and then the effectiveness of the system is tested by presenting the test data to it. The paper is organized as follows. Section 2 explains the multiclass classification in SVM and Section 3 describes the hybrid kernel function and proposed kernel function. Section 4 discusses the experimental results and Section 5 concludes the work.

## 2. MULTICLASS CLASSIFICATION (Vojtech Franc and Václav Hlavác, 2009)

Multiclass SVM can be solved by combining the binary classification decision functions. Multiclass SVM is of two types namely **One against One** (OAO) and **One against All** (OAA) decomposition. The OAA decomposition transforms the multiclass problem into a series of  $c$  binary subtasks that can be trained by the binary SVM. Let the training set  $T_{XY}^y = \{(x_1, y_1), \dots, (x_l, y_l)\}$  contain the modified hidden states defined as

$$y_i = \begin{cases} 1 & \text{for } y = y_i, \\ 2 & \text{for } y \neq y_i \end{cases} \quad (1)$$

The discriminant functions

$$f_y(x) = \langle \alpha_y \bullet K_s(x) \rangle + b_y, \quad y \in Y, \quad (2)$$

are trained by the binary SVM solver from the set  $T_{XY}^y, y \in Y$

The OAO decomposition transforms the multi-class problem into a series of  $g = c(c-1)/2$  binary subtasks that can be trained by the binary SVM. Let the training set  $T_{XY}^y = \{(x_1^j, y_1^j), \dots, (x_{l_j}^j, y_{l_j}^j)\}$  contain the training vectors  $x_i \in I = \{i: y_i = y^1 \vee y_i = y^2\}$  and the modified hidden states defined as

$$y_i = \begin{cases} 1 & \text{for } y_j^1 = y_i, \\ 2 & \text{for } y_j^2 \neq y_i, \end{cases} \quad i \in I^j \quad (3)$$

The training set  $T_{XY}^j, j = 1, 2, \dots, g$  is constructed for all  $g=c(c-1)/2$  combinations of classes  $y_j^1 \in Y \ \& \ y_j^2 \in Y \setminus \mathcal{R}_j$ . The binary SVM rules  $q_j, j = 1, \dots, g$  are trained on the data  $T_{XY}^j$ .

## 3. KERNEL FUNCTION

From the empirical analysis, the two kernel functions that have given better results for the datasets when compared to other kernels are selected to construct a hybrid kernel. The chosen kernel functions from the experimental analysis are radial basis kernel and polynomial kernel. Therefore, a kernel function is designed by hybridizing the translation invariant and rotation invariant kernels i.e. RBF and Polynomial. The hybrid kernel is given by the following equation

$$K(x, y) = \lambda K_1 + (1 - \lambda) K_2 \quad (4)$$

where  $K_1$  is RBF kernel,  $K_2$  is polynomial kernel and  $\lambda$  is a box constraint that control and balance these two kernels.

From an extensive analysis (Song et al. 2012), admissible kernel functions in Banach space are designed either based on exponential distribution or Brownian bridge distribution. But, these distributions have been devoted to the class of processes with stochastic integrals, geometry of Banach space, stationary processes and independent increments called as Lévy processes (Applebaum, 2007; FAJ Ueltzhöfer, 2012). Lévy process is an origin of the mainly used distributions in Banach space since lévy is strongly interrelated with Banach space geometry e.g. Brownian process and poisson process. Thus, a novel kernel is constructed in RKBS using lévy distribution (Sangeetha and Kalpana, 2013).

Novel kernel function using lévy distribution is defined by the following equation

$$K(s, t) = \frac{e^{-c|s-t|}}{|s-t|} \quad (5)$$

where  $C$  is regularization parameter and  $s, t \in \mathbb{R}$ .

**Table 1. Multiclass Datasets from UCI Machine Learning Repository**

Data Sets	Instances	Attribute	Class Distribution
Iris	150	4	3
Glass	214	10	7
E-Coli	336	8	8
Wine	178	13	3
Balance Scale	625	4	3
Lenses	24	4	3
Pentagon	99	2	5

#### 4. EXPERIMENTAL RESULTS

The experimentation for multiclass datasets using kernel functions are carried out using MATLAB and C. The multiclass datasets are taken from the UCI machine learning repository which is detailed in Table 1. Before applying the classification process, the actual datasets are preprocessed and its dimensionality is reduced. Here, the proposed and hybrid kernel functions are evaluated using the performance metrics like number of Support Vectors (SV), support vector percentage, training error, classification error and time. They are given by the following equations

##### Number of Support Vectors

$$\text{Vectors } \mathbf{x}_i \text{ for which } \mathbf{y}_i (\mathbf{w} \cdot \mathbf{x}_i + \mathbf{b}) = 1, \quad (6)$$

##### Support Vector Percentage

$$SV\% = \frac{\text{Number of support vectors}}{\text{Number of training data}} * 100 \quad (7)$$

##### Training Error

$$Err = \frac{1}{N} \sum_{i=1}^N L(y_i, \hat{f}(x_i)) \quad (8)$$

##### Testing Error

$$Err = E[L(y, \hat{f}(x))] \quad (9)$$

##### Accuracy

$$\text{Accuracy} = \frac{\text{Total number of correctly classified samples}}{\text{Total number of samples}} \quad (10)$$

where  $\mathbf{x}$  is an input data,  $\mathbf{y}$  is the class label,  $\mathbf{w}$  is weight vector,  $\mathbf{b}$  is bias value,  $\hat{f}(x)$  is predicted class label,  $N$  is number of samples,  $L$  is loss function to estimate an error.

Classification time and computational complexity for the SVM multiclass classifier depend on the number of support vectors required for the classification process. As the number of support vector increases, it leads to increase in computational requirements such as addition, multiplication and floating point. In SVM, for multiclass classification the required memory to store the support vectors is directly proportional to the number of support vectors. Hence, the number of support vectors must be reduced to speed up the classification and to minimize the computational and hardware resources required for the classification process.

**Table 2. Performance of Kernel Functions for Multiclass datasets using One Against One SVM**

Datasets	Proposed Kernel				Hybrid Kernel			
	SV	TR	TE	Time	SV	TR	TE	Time
Iris	31	0.0166	0.1	0.078	35	0.025	0.2	0.046
Glass	54	0.142	0.036	0.375	80	0.3969	0.45	0.08
E-Coli	98	0.241	0.375	0.02	134	0.324	0.497	0.54
Wine	52	0.3194	0.0294	1.015	93	0.0277	0.2058	1.093
Scale	95	0.08	0.296	11.15	124	0.326	0.48	17.81
Lenses	10	0.45	0.1666	0.031	14	0.5	0.3333	0.01
Pentagon	44	0.21	0.25	0.078	73	0.1392	0.25	0.062

**Table 3. Percentage of Accuracy and Support Vectors for Multiclass datasets using One Against One SVM**

Datasets	Proposed Kernel		Hybrid Kernel	
	SV %	Accuracy	SV%	Accuracy
Iris	25.8	90	29.16	80
Glass	41.2	63.4	61.06	55
E-Coli	36.56	62.5	50	50.28
Wine	36.11	97.05	64.5	79.41
Scale	19	70.4	24.8	52
Lenses	55.55	83.33	77.77	66.66
Pentagon	55.69	75	92.4	75
Average	<b>38.55</b>	<b>77.38</b>	57.09	65.47

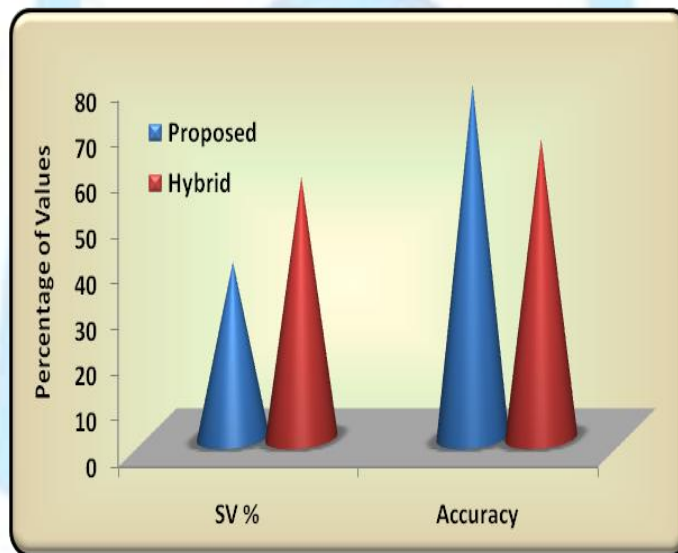
**Table 4. Performance of Kernel Functions for Multiclass datasets using One against All SVM**

Datasets	Proposed Kernel				Hybrid Kernel			
	SV	TR	TE	Time	SV	TR	TE	Time
Iris	22	0.0166	0.1	0.125	20	0.025	0.1333	0.062
Glass	78	0.396	0.42	0.25	90	0.462	0.5	0.75
E-Coli	95	0.224	0.386	0.05	140	0.371	0.477	0.24
Wine	37	0.1458	0.3235	16.25	95	0.0277	0.428	1.39
Scale	110	0.302	0.315	15.6	120	0.56	0.44	18.1
Lenses	8	0.32	0.144	0.015	11	0.47	0.304	0.06
Pentagon	45	0.0126	0.0	0.031	41	0.3924	0.55	0.078

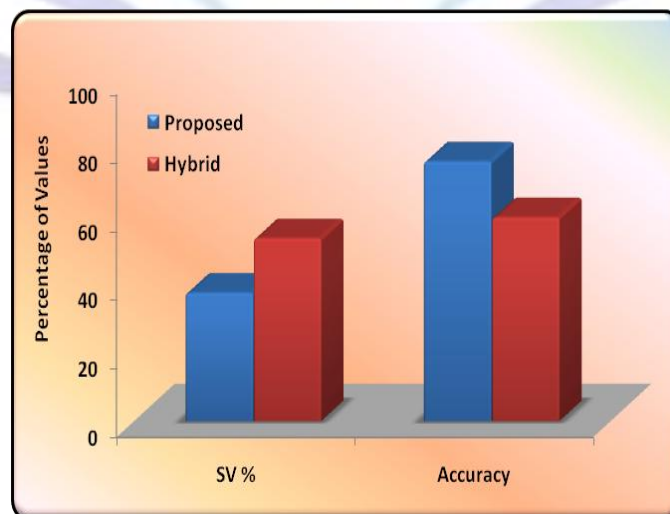
Here, the proposed kernel function and hybrid kernel function are executed with multiclass benchmark datasets. To overcome the demerits of predefined kernel functions in Hilbert space, a hybrid kernel function in Hilbert space is designed. But from the observed results, it is evident that there are some limitations in Hilbert space. So, the motivation for constructing a novel and admissible kernel functions in Banach space has emerged.

**Table 5. Percentage of Accuracy and Support Vectors for Multiclass datasets using One Against All SVM**

Datasets	Proposed Kernel		Hybrid Kernel	
	SV %	Accuracy	SV%	Accuracy
Iris	18.33	90	16.66	86.66
Glass	59.5	58	68.7	50
E-Coli	35.44	61.4	85	52.23
Wine	25.69	67.64	65.97	57.2
Scale	22	68.5	24	56
Lenses	44.44	85.6	61.6	69.6
Pentagon	56.96	100	51.89	45
<b>Average</b>	<b>37.48</b>	<b>75.87</b>	53.4	59.53



**Fig 1: Overall percentage of Accuracy and Support Vectors for Multiclass datasets using One against One SVM**



**Fig 2: Overall percentage of Accuracy and Support Vectors for Multiclass datasets using One against All SVM**



The experiments for the proposed kernel and hybrid kernel in OAO and OAA SVM are carried out using multiclass datasets. The results for OAO and OAA SVM are given in the Tables 2-5. The overall accuracy for multiclass datasets using OAO SVM and OAA SVM are 77.38 % and 75.85 %, which are represented in the Figures 1 and 2. From the observed results, it is identified that the proposed kernel function outperforms the hybrid kernel function by increasing the classification accuracy and minimizing the training error as well as the classification error.

#### 4. CONCLUSION

The major drawback of the kernel based learning algorithms is the absence of predefined hypothesis to choose a feasible kernel function for specific applications. To overcome this challenge, a novel kernel function that satisfies an admissible kernel property in Banach space is designed, which it can be applied to various domains. The proposed admissible kernel function in RKBS overcomes the three main disadvantages of the existing kernels i.e. it improves the generalization ability of a classifier, enhances the performance of a classifier and tunes the parameters to attain a best fit solution. From the extensive experimental results, it is shown that the proposed admissible kernel function improves the performance of accuracy by 11.91% for OAO multiclass SVM and 16.34 % for OAA multiclass SVM.

#### REFERENCES

- [1] Sangeetha,R. and Kalpana,B. 2010.Supervised Classification Techniques - A Review, National Conference on Information and Communication Technology - NCICT-2010.
- [2] . Dutt,V., Chadhury,V. and Khan,T. 2011. Different Approaches in Pattern Recognition, Computer Science and Engineering,Volume 1,Issue2,pp. 32-35.
- [3] . Sharma, A.M. and Kishor, R.R. 2007. Pattern Recognition: Different Available Approaches, Proceedings of National Conference on Challenges & Opportunities in Information Technology, RIMT-IET.
- [4] . Franc, V., and Hlavác. V. (2008). Statistical Pattern Recognition toolbox for Matlab.
- [5] . Song, G., Zhang, H., and Hickernell, F. J. 2012. Reproducing kernel Banach spaces with the  $\ell_1$  norm. Applied and Computational Harmonic Analysis.
- [6] . Applebaum, D. 2007. Lévy processes and stochastic integrals in Banach spaces. Probability and Mathematical Statistics-Wroclaw University, 27,75-88. 3.
- [7] Sangeetha,R. and Kalpana,B. 2013.A Mathematical Framework for an Admissible Kernel Function in Support Vector Machines using Lévy Distribution. (Communicated)