

Comparison of Support Vector Machines and Linear Discriminant Analysis for Indian Industries

R. Madhanagopal^{1*}, R.C. Avinaash² and K. Karthick³

{¹Assistant Professor, ^{2,3}Post Graduate Students} Department of Statistics, Madras Christian College, Chennai, Tamil Nadu, INDIA.

Corresponding Addresses:

¹madhan.stat@gmail.com, ²avinaash2008@gmail.com, ³k.karthick14@gmail.com

Research Article

Abstract: Support vector machines (SVM) methods have become increasingly popular tools for data mining tasks *viz.*, classification, regression and novelty detection. The present paper deals with classification of Indian industries using SVM. Industries stable for one month period in NIFTY was selected, of which 50 companies in NIFTY, 32 were found to be stable. Twenty eight key financial ratios of these companies were taken for a period of five financial years (April 2007 to March 2012). Fuzzy clustering and SVM were used to explore the financial data. Principal component analysis (PCA) was applied and it reduced the twenty eight financial ratios into seven components. Thereafter, fuzzy clustering was performed on scores of PCA and was formed into two groups which were categorized into high and low performing industries based on their mean values. SVM was used as a classifier of the industries and it was compared with well known and old classification technique, Linear discriminant analysis (LDA). The classification accuracy in training and testing data set for SVM was 97.32% and 100 % whereas for LDA it was 87.29 and 93.75% respectively. Therefore, the present study concludes that SVM performed better than LDA in the classification of industries.

Key words: Financial Ratio, Classification, Support Vector Machines, Fuzzy Clustering, Principal Component Analysis.

1. Introduction

Livelihood of the people changes due to the development of economy in the country and especially in developing countries, industries plays a vital role in the development of the country's economy. According to [6] one-third of the population of the world lived in poverty in 1981, whereas the share was 18% in 2001. This huge decline was due to the economic development in India and China. Indian Gross domestic product (GDP) increases from 3.9 in 2001 to 7.2 in 2011. In this, the contribution of industries was 20.16, second to services sector contribution of 65.22%. Rise of industries in a country boost the employment opportunity, income and saving, economic scale and farm productivity. On the other hand it declines the poverty, crime, society imbalance, etc. Due to globalization and liberalization in Indian government policy, many new industries from inside and outside the county has emerged in recent years but it increases the competitive nature to survive which resulted the industries to monitor their performances regularly which is not an easy task. One of the ways to supervise them is by financial ratios and therefore evaluation on the performances of the industries is inevitable. A

support vector machine (SVM) is a training algorithm for learning classification and regression rules from data ([23], [36], [7], [21]). SVM is applied successfully in many areas such as face detection ([12], [2], [29], [14]), image classification ([38], [8]), object recognition ([28], [37], [26]), hand written / digital recognition ([13], [3], [22], [1], [16]), speaker speech recognition ([5], [33], [24]), gender classification ([17], [35], [10], [27]), text classification ([32], [18]) etc.. Several recent studies have reported that SVM is capable of rendering higher performance in terms of classification accuracy than other data classification ([11], [34]). Therefore, the present paper deals with usage of SVM as a tool to classify the Indian industries then to check whether SVM classifies Indian industries better than Linear Discriminant Analysis (LDA). Rest of the paper is organized as follows; section two deals with the selection of samples, data description and methodology used in the present study. Brief introduction of data analysis techniques *viz.*, principal component analysis for data reduction, fuzzy clustering for clustering, the industries into homogenous groups and SVM classifier are described in section three. Findings and discussions of the results are presented in section four and conclusion in section five.

2. Data and methodology

2.1 Sample selection and data description

The study was analytical in nature and the present study uses the latest available published secondary data starting from April 2007 to March 2012. The units of analysis include 50 industries that are listed on Nifty. Thirty two industries were filtered based on the following criteria. i) The industries must be listed on Nifty. ii) The industry must be stable in the list of Nifty for period of a month (1st - 30th September, 2012). iii) The data of variables for industries must be available for the period of study. iv) Financial service based industry *viz.*, banks, financial intuitions etc., were excluded. Financial ratios provide a quick and relatively simple means of examining the financial condition of an industry since it is of very good help when comparing the financial health of different businesses [19]. Therefore, to identify the financial performance of Indian industries, financial ratios of the industries were

used. By carefully examining the previous studies 28 most important financial ratios were selected. Financial ratios were obtained from income statements, balance sheet, cash flow data sheet, etc., of the industries. For this study, these ratios were extracted from money control web page (www.moneycontrol.com) and the

needed financial ratios were calculated. These ratios were selected to assess profitability, investment values, liquidity, solvency, debt coverage, management efficiency, profit and loss. The variables (financial ratios) used and its codes are shown in Table 2.1.1.

Table 2.1.1. List of financial ratios and its codes

Financial Ratio	Codes	Financial Ratio	Codes
Operating Profit Per Share	OPPS	Interest Cover	INC
Net Operating Profit Per Share	NOPPS	Total Debt to Owners Fund	TDTOF
Operating Profit Margin	OPM	Financial Charges Coverage Ratio	FCCR
Gross Profit Margin	GPM	Inventory Turnover Ratio	ITR
Cash Profit Margin	CPM	Debtors Turnover Ratio	DTR
Net Profit Margin	NPM	Total Assets Turnover Ratio	TATR
Return On Capital Employed	ROCE	Number of Days In Working Capital	NDWC
Return On Net Worth	RONW	Material Cost Composition	MCC
Return on Assets Including Revaluations	ROAIR	Selling Distribution Cost Composition	SDCC
Return on Long Term Funds	ROLTF	Expenses Total Sales	ETS
Current Ratio	CUR	Dividend Payout Ratio Net Profit	DPRNP
Quick Ratio	QUR	Dividend Payout Ratio Cash Profit	DPRCP
Debt Equity Ratio	DER	Earning Retention Ratio	ERR
Long Term Debt Equity Ratio	LTDER	Cash Earning Retention Ratio	CERR

2.2 Methodology

Firstly, 28 financial ratios which were used in the present study are normalized [0, 1] using the formula.

$$y_i = [(x_i - \min_j(x_j)) / [\max_j(x_j) - \min_j(x_j)]]$$

For the normalized variables, principal component analysis was applied to avoid the influence of correlation among the variables and, the principal scores obtained were used as the variables for the rest of this study. Secondly, fuzzy clustering was applied to the principal scores to group the industries. SVM classifier with different kernels and Linear discriminant analysis (LDA) were performed to compare classification efficiency of the classification tools.

3. Data analysis techniques

3.1 Principal component analysis (PCA)

The technique of Principal component analysis (PCA) was first described by [25] and is one of the simplest techniques of the multivariate methods. The main objectives of PCA are to identify new meaningful underlying variables and discover or reduce the dimensionality of the data set. PCA involves a mathematical procedure that transforms a number of correlation variables into a number of uncorrelated variables called principal components. The lack of correlation is most important and is a useful property because the uncorrelated variables are measuring different dimensions in the data. These uncorrelated variables are ordered based on its variation i.e., largest amount of variation displayed first, followed by second largest amount [20].

Steps involved in construction of PCA

Let us consider p variables say $x_1, x_2, x_3, \dots, x_p$ for the study

- 1) First normalize the data.
- 2) Calculate the correlation matrix C.

- 3) Find the eigen values $\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_p$ and the corresponding eigen vectors $a_1, a_2, a_3, \dots, a_p$. The coefficients of the i^{th} principal component are the given by a_i while λ_i is its variance.
- 4) Discard any components that only account for a small proportion of the variation in the data.

3.2 Fuzzy Clustering (FC)

Clustering is a division of data into groups of similar objects. There are different clustering techniques available of which most of them belongs to hard/conventional clustering (every data points belongs to one unique cluster). In many situations, clusters are not well separated as a result, the data point may or may not belong to a particular cluster. Fuzzy clustering differs from other clustering techniques. In fuzzy clustering, object may belong to more than one cluster with varying degrees of membership. Fuzzy cluster analysis has its origins by [31] and the fuzzy clustering approach is based on the fuzzy set theory proposed by [39]. There are different approaches for fuzzy clustering. Fuzzy k – means, the earliest methods was proposed by [9] and [3] wherein Fuzzy k-means is a generalization of the crisp k-means clustering. In the present paper, FUNNY algorithm was used attributed to [15]. The objective function in which the dissimilarity or distance measure is a L^1 expression (not squared) is used, which makes it more robust than fuzzy k- means [30]. FUNNY aims at the minimization of the following objective function.

$$Objective\ Function\ C = \sum_{r=1}^m \frac{\sum_{i,j=1}^n u_{ir}^2 u_{jr}^2 d(i,j)}{2 \sum_{j=1}^n u_{jr}^2}$$

Where $d(i,j)$ represent the given distance (or dissimilarities) between objects i and j .

u_{ir} = the unknown membership of object i to cluster n .

m = the number of clusters that the fuzzy clustering solution will have.

n = the number of observation in the data set.

The membership functions are subject to the constraints.

- i) $u_{ir} \geq 0 \forall i = 1, \dots, n; r = 1, \dots, m$
- ii) $\sum_r u_{ir} = 1$ for $i = 1, \dots, n$

These constraints imply that membership cannot be negative and that each object has a certain total membership distributed over different clusters. By convention, the above mentioned total membership is normalized to 1. The solution to above mentioned optimization problem is iterative.

To test goodness of fit for fuzzy clustering, two methods were used.

i) [15] proposed the silhouette statistic for assessing clusters and estimating the optimal number. For observation i , let $a(i)$ be the average distance to other points in its cluster, and $b(i)$ the average distance to points in the nearest cluster besides its own nearest is defined by the cluster minimizing this average distance followed by silhouette statistic defined by

$$S(i) = \frac{b(i) - a(i)}{\max\{a(i) - b(i)\}}$$

A point is well clustered if $S(i)$ is large. [15] Proposed to choose the optimal number of clusters \hat{k} as the value maximizing the average $S(i)$ over the data set.

ii) Dunn's partition coefficient (DPC) was introduced to identify the compact and well separate cluster. Each object in the fuzzy cluster has a membership and if it is 1 in one cluster and 0 for all other cluster then cluster is a hard entirely. This coefficient can be use to know whether cluster is hard or fuzzy nature. DPC is computed by the following formula for which values lie between $(\frac{1}{k}, 1)$.

$$D_m = \sum_{i=1}^n \sum_{r=1}^m \frac{u_{ir}^2}{n}$$

Where

- u = final membership matrix,
- u_{ir} = membership value of the i^{th} element to the r^{th} cluster.
- n = number of observation and
- m = number of cluster.

This can be normalized and normal version of the Dunn's partition coefficient (DPC) was obtained from the formula given below. The values of normalized DPC always lie in range (0, 1). For a good clustering solution, this value should be high.

$$ND_m = \frac{D_m - \frac{1}{m}}{1 - \frac{1}{m}} = \frac{m D_m(u) - 1}{m - 1}$$

3.3. Support Vector Machines (SVM)

SVM is an emerging machine learning technique in the field of data mining. It a novel learning machine introduced first by [36]. A brief mathematical background of SVM is given below.

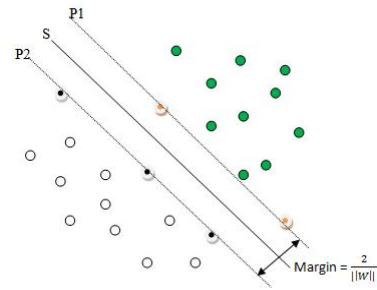


Figure 3.1. Optimal Separating Hyper plane for linear

In figure 3.1 there are two types of dots one is green in color and the other is white representing two kinds of samples. S is the separating line and P1 and P2 are the closest lines parallel to the separating line of the two class sample vectors. The distance between P1 and P2 is called the margin.

For linear case,

Given a training set belong to two separate classes say (x_i, y_i) , where $i = 1, 2, 3, \dots, n, x \in R^n, y \in \{-1, 1\}$ with a hyperplane,

$$(\omega \cdot x) + b = 0 \quad (3.1)$$

equ (3.1) is the separating hyperplane equation. The training class (x_i, y_i) should satisfy

$$y_i[(\omega \cdot x) + b] \geq 1 \text{ where } i = 1, 2, 3, \dots, n \quad (3.2)$$

The distance of point x to the hyperplane (ω, b) is $(\omega, b; x) = \frac{|(\omega \cdot x') + b|}{\|\omega\|}$. The optimal hyperplane is given by maximizing the margin d , subject to equation (3.2). The margin can be given by $\rho(\omega, b) = \frac{2}{\|\omega\|}$.

Hence the hyperplane that optimally separates the data is the one that minimizes

$$\phi(\omega) = \frac{\|\omega\|}{2} \quad (3.3)$$

The Lagrange function of (3.3) under constraints (3.2) is,

$$\phi(\omega, b; a) = \frac{1}{2} \|\omega\|^2 - \sum_{i=1}^l \alpha_i (y^i [(\omega \cdot x^i) + b] - 1) \quad (3.4)$$

The optimal classification function, if solved, is

$$f(x) = \text{sgn}(\omega^* \cdot x) + b^*$$

For non linear case,

Let $\Phi: R^d \rightarrow H$ be a non linear map which transforms x from the input space into the feature space H. A kernel is a function $k, \forall x_1, x_2 \in R^d, K(x_1, x_2) = \langle \Phi(x_1), \Phi(x_2) \rangle$. Then, for the nonlinear classification, to determine the optimal hyperplane equals to solve the following constrained quadratic optimization problem.

Maximize the objective function

$$Q(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j) \quad (3.5)$$

Subject to the constrains

$$\sum_{j=1}^n \alpha_j y_j = 0, \alpha_i \geq 0, i = 1, 2, \dots, n \quad (3.6)$$

where α_i is the Lagrange multiplier for each sample.

The corresponding separating function is

$$f(x) = \text{sgn}(\sum_{i=1}^n \alpha_i^* y_i k(x_i, x) + b^*) \quad (3.7)$$

This is the so-called SVM.

Kernels used in this paper are as follows

- LINEAR $K(x_i, x_j) = x_i^T x_j$
- POLYNOMIAL $K(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \gamma > 0$
- RBF $K(x_i, x_j) = \exp(-\gamma |x_i - x_j|^2) \gamma > 0$
- SIGMOID $K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r)$

4. Findings and discussion

4.1 Results of principal component analysis (PCA)

The application of the PCA is to reduce the dimensionality of the data set and also to avoid the influence of correlation among the variables. As discussed earlier raw data set of 28 financial ratios of the Indian industries are normalized and robust principal component analysis was performed. This was done by using R-2.15.1 software. This produce eigen vector (e_i) and eigen value (λ_i). The eigen vector were ordered so that $\lambda_i < \lambda_j$. Thus the lower order eigenvectors encode the majority of the variances.

Table 4.1.1. Eigen values, percentage and Cumulative Percentage of total variance

Components	Eigen Value	% Variance	Cumulative % variance
PC2	6.6515	23.7555	23.7555
PC1	5.7424	20.5085	44.264
PC7	3.5156	12.5557	56.8197
PC3	2.9115	10.398	67.2177
PC5	1.4938	5.3351	72.5528
PC6	1.4043	5.0153	77.5681
PC4	1.1610	4.1465	81.7146

From table 4.1.1, by taking 1 for the eigen value as the cutoff point we have selected the seven components, the selected components approximately represents 82% of the variance structure of the raw data. Figure 4.1.1 Shows the 3D scatterplot of first 3 principal components as axes.

3 D SCATTERPLOT FOR FIRST 3 PRINCIPAL COMPONENTS

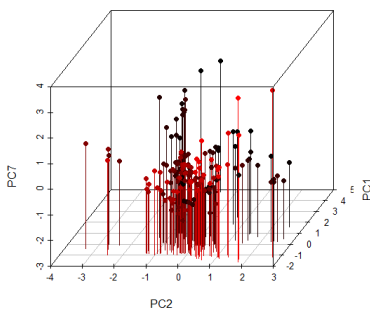


Figure 4. 1.1 3D plot of first three Principal Components

4.2 Results of fuzzy clustering

Fuzzy clustering was applied to the principal scores extracted from PCA to group the industries. Squared Euclidean distance is used as the metric of dissimilarity in the data. Squared Euclidean distance between two points, a and b , with k dimensions is calculated as (4.2.1). When taken square root of equation (8), it becomes Euclidean distance.

$$D = \sum_{j=1}^k (a_j - b_j)^2 \quad (4.2.1)$$

Fuzzy clustering was carried out next using the [15] algorithm used in FUNNY package. R-2.15 software package is used for analyzing the data. The process of selecting the best clusters is subjective in nature. Several clustering solutions are generated using different number of clusters, different values of fuzzier and different dissimilarities. Based on various indices such as the silhouette and Dunn's partition index for fuzzier 1.25 and square euclidean dissimilarity, 2 clusters are chosen. The final solutions of various indices and silhouette using fuzzier at 1.25 are presented below table 4.2.1.

Tabl 4.2.1. Summary of number of Fuzzy Clustering and its Silhouette and Dunn's Index (m = 1.25, Metric = Square Euclidean dissimilarity)

Number of clusters	S(i)	Dunn's Partition Index	Normalized Dunn's Partition Index
2	0.31	0.79253	0.75103
3	0.30	0.77579	0.71974
4	0.26	0.74178	0.65571
5	0.16	0.71312	0.56969
6	0.25	0.74510	0.49020

For the best clustering, silhouette plot and fuzzy clustering plot are shown below in Figure 4.2.1 and 4.2.2.

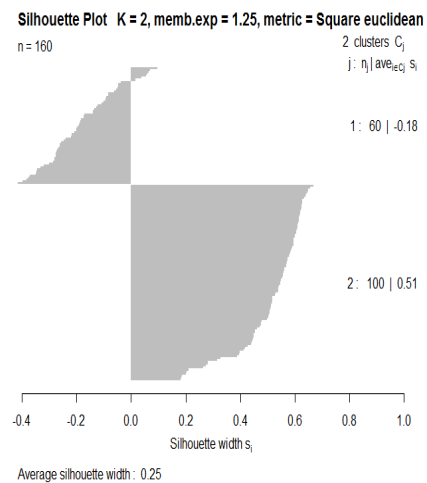


Figure 4.2.1. Silhouette plot for K = 2, Fuzzier = 1.25 and Square Euclidean dissimilarity

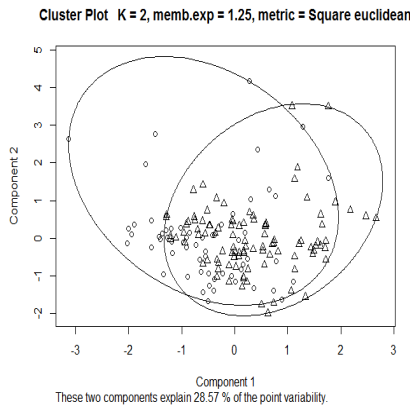


Figure 4.2.2 Fuzzy Clustering plot for $K = 2$, Fuzzier = 1.25 and Square Euclidean dissimilarity

On the bases of membership coefficients, 160 industries are grouped into two groups. Based on the mean values, these two groups are categories as high and low performing industries. 60 industries belong to high performing and 100 industries in low performing group.

4.3 Comparison of SVM and LDA

The whole data set used in fuzzy clustering was divided into two sets viz., training and testing data. It was done in standard spreadsheet software. Approximately, 70% of the industries were assigned to the training sample. The training data set resulted in 112 industries and the remaining 48 industries were assigned to the test sample. SVM classifier was performed with different kernels for training and testing data set. The performance of SVM was compared with a well know classification method LDA. From Table 4.3.1 and 4.3.2 it was observed that accuracy and error rate of SVM for training and testing was (97.321 and 100.00%) and (2.679 and 0.000%) whereas for LDA (89.286 and 93.750%) and (10.714 and 6.250%) respectively.

Table 4.3.1. Confusion matrix of training and testing for SVM (linear)

		Training			Testing			
		Predicated by Model			Predicated by Model			
		HIGH	LOW		HIGH	LOW		
Actual	HIGH	42	2	44	HIGH	16	0	16
	LOW	1	67	68	LOW	0	32	32
		43	69	112		16	32	48
	AR	97.321			100.00			
	ER	02.679			0.000			

AR – Accuracy rate, ER – Error rate

Table 4.3.2. Confusion matrix of training and testing for LDA

		Training			Testing			
		Predicated by Model			Predicated by Model			
		HIGH	LOW		HIGH	LOW		
Actual	HIGH	35	9	44	HIGH	15	1	16
	LOW	3	65	68	LOW	2	30	32
		38	74	112		17	31	48
	AR	89.286			93.750			
	ER	10.714			6.250			

Table 4.3.3 shows the accuracy and error rate of SVM for different kernels with gamma 0.142 and cost 10. Error rate of SVM (RBF) was 0 and 2.083%. Similarly for SVM (Polynomial and Sigmoid) it was 1.786 and 13.793% and 2.083 and 9.231% for training and testing data sets indicating SVM with RBF kernel performed well in classification than other kernels and LDA.

Table 4.3.3 Results of classification efficiency of SVM (RBF, Polynomial and Sigmoid)

Kernels	SVM		
	RBF	Polynomial	Sigmoid
Gamma	0.142	0.142	0.142
Cost	10	10	10
TRAINING			
AR	100.000	98.214	86.207
ER	0.000	1.786	13.793
TESTING			
AR	97.917	97.917	90.769
ER	2.083	2.083	9.231

AR – Accuracy rate, ER – Error rate

Comparison of accuracy rate for SVM with different kernels and LDA for training and testing are shown in Figure 4.3.1. Overall, the result shows other than sigmoid kernel, SVM outperformed well than LDA. Therefore, it can be concluded that performance of SVM in classifying the Indian industries is better than LDA.

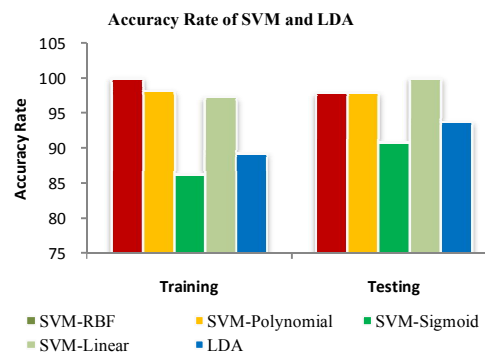


Figure 4.3.1. Comparison of accuracy rate for SVM with different kernels and LDA for training and testing.

Conclusion

The present study was evaluated to check whether SVM classifies Indian industries better than Linear Discriminant Analysis (LDA). Four different kernels for SVM viz., RBF, Polynomial, Sigmoid and Linear were used and compared with the classification results of each other and also with LDA. The result showed that error rate of SVM (Linear) was 2.679 whereas LDA was 10.714% for training and 0 and 6.250% for testing data sets. The difference in error rate between SVM (Linear) and LDA was 8.035 and 6.250%. Therefore, it can be concluded that SVM (Linear) performs better than LDA. Similarly for SVM (RBF, Polynomial and Sigmoid) error rate was 0, 1.786 and 13.793% for training and 2.083, 2.083 and 9.231% for testing data sets indicating that SVM (RBF) performed well in classification than other kernels and LDA. Overall, the result shows other than sigmoid kernel, SVM outperformed well than LDA and it can be concluded that performance of SVM in classifying the Indian industries is better than LDA.

Acknowledgement

The authors would like to thank Dr. R. Chandrasekaran, Professor and Head (retired), Department of Statistics, Madras Christian College, Chennai, Tamil Nadu, India for his valuable suggestions and Dr. Samtennyson, Department of Zoology, Madras Christian College, Chennai, Tamil Nadu, India for his timely help.

References

1. Aggarwal, A. Rani, R. and Dhir, R, "Recognition of devanagari handwritten numerals using gradient features and SVM", international journal of computer applications, Vol. 48, No.8, pp.39-44, 2012.
2. Ai, H. Liang, L. and Xu, G, "Face detection based on template matching and support vector machines", proceedings of international conference on image processing, pp. 1006-1009, 2001.
3. Bezdek, J.C, "Numerical Taxonomy with fuzzy sets", journal of mathematical biology, Vol.1, pp.57-71, (1974).
4. Arora, S. Bhattacharjee, D. Nasipuri, M. Malik, L. Kundu, M and Basu, D. K, "performance comparison of SVM and ANN for handwritten devnagari character recognition", international journal of computer science issues, Vol. 7, No.3, pp.18-26, 2010.
5. Chavhan, Y. Dhore, M. L., Yesaware, P, "Speech emotion recognition using support vector machine", international journal of computer applications, Vol.1, pp.6-9, 2010.
6. Chen, S and Ravallion, M, "How have the world's poorest fared since the early 1980s?" ,world bank research observation, Vol. 19, No. 2, pp. 141 – 170, 2004.
7. Cristianini N, Taylor, J. S, "An introduction to support vector machines", cambridge university press, cambridge, NewYork 2000.
8. Dhasal, P. Shrivastava, S. S, Gupta, H. and Kumar, P, "An optimized feature selection for image classification based on SVM-ACO", international journal of advanced computer research, Vol.2, No. 5, pp.123 – 128, 2012.
9. Dunn, J. C, "A fuzzy relative of the ISODATA process and its use in detecting compact well- separated clusters", Journal of Cybernetics, Vol.3, pp.32-57, 1974.
10. Gaikwad, S. Gawali, B and Mehrotra, S. C, "Gender identification using SVM with combination of MFCC", advance in computational research, Vol. 4, issue 1, pp.69-73, 2012.
11. Gokcen, I and Peng, J, "Comparing linear discriminant analysis and support vector machines", advance in information systems, lecture notes in computer science, Vol. 2457, pp. 104-113, 2002.
12. Guodong, G, S. Li, and C. Kapluk, "Face recognition by support vector machines", in proceedings IEEE international conference on automatic face and gesture recognition, pp.196–201, 2000.
13. Hamdi, R. Bouchareb, F. and Bedda, M, "Handwritten Arabic character recognition based on SVM classifier", information and communication technologies: from theory to applications, 3rd international conference on sep. 7-11, Damascus, 2008.
14. Jia, H. and Martinez. A. M, "Face recognition with occlusions in the training and testing sets", Proceedings of the IEEE international conference on automatic face and gesture recognition, 2008.
15. Kaufman, L. and Rousseeuw, P. J, "Finding groups in data: An introduction to cluster analysis", John Wiley and Sons, New York. 1990.
16. Kumar, P. Sharma, N. and Rana, "A. Hand written character recognition using different kernel based SVM classifier and MLP neural network (A COMPARISON)", international journal of computer application, Vol. 53, No.11, pp.25-31, 2012.
17. Lian, H.C. and Lu, B. L, "Multi-view gender classification using multi-resolution local binary patterns and support vector machines", international journal of neural system, Vol.17, No.16, pp.479 – 487, 2007.
18. Lewis D. L., Yang Y., Rose T. G., Li, F, "A new benchmark collection for text categorization research", journal of machine learning research, Vol.5, pp.361-397, 2004.
19. Mahmoud, O. H, "A multivariate model for predicting the efficiency of financial performance for property and liability Egyptian insurance companies", casualty actuarial society, discussion paper, 2008.
20. Manly, B. F. J, "Multivariate statistical method a primer", chapman and hall, New York, 1986.
21. Muller K. R, Mika S, "An introduction to kernel-based learning algorithms", IEEE transactions on neural networks, Vol.12, No.2, pp.181– 201, 2001.
22. Nasien, D. Haron, H and Yuhaniz, S. S, "Support vector machine (SVM) for English handwritten character recognition", computer engineering and applications (ICCEA), second international conference, Vol.1, pp.249-252, 2010.
23. Osuna, E., R. Freund and F. Girosit, "Training support vector machines: an application to face detection", proceedings of IEEE computer society conference on computer vision and pattern recognition, june 17-19, Puerto Rico, pp: 130-136, 1997.
24. Pan, Y. Shen, P. and Shen, L, "Speech emotion recognition using support vector machine", international journal of smart home, Vol. 6 No.2, pp. 101 – 108, 2012.

25. Person, K, "On lines and planes of closest fit to a system of points in space", philosophical magazine, Vol.2, pp 557- 572, 1901.
26. Petropoulos, P. G. Kalaitzidis, C. and Vadrevu, K. P, "Support vector machines and object-based classification for obtaining land-use / cover cartography from hyperion hyperspectral imagery", computers and geosciences, vol. 41, pp. 99-107, 2012.
27. Ponnarasi, S. S and Rajaram, M, "Gender classification system derived from fingerprint minutiae extraction". IJCA proceedings on international conference in recent trends in computational methods, communication and controls (ICON3C 2012) pp.1-6, 2012.
28. Pontil, M. and Verri, A, "Properties of support vector machines", neural computation, Vol. 10, No.4, pp. 955–974, 1998.
29. Romdhani, S. Torr, P. and Scholkopf, B, "Efficient face detection by a cascaded support-vector machine expansion", royal society of London proceedings, series A, Vol. 460, pp. 3283–3297, 2004.
30. Rousseeuw, P, "Fuzzy clustering at the intersection", Technometrics: A journal of statistics for the physical, chemical and engineering sciences, Vol.37, pp. 283-286, 1995.
31. Ruspini, E. H, "A new approach to clustering, information and control, Vol.15, pp. 22-32, 1969.
32. Sebastiani, F, "Machine learning in automated text categorization". ACM computing surveys vol.34, No.1, pp.1–47, 2002.
33. Shen, P. Changjun, Z and Chen, X, "Automatic speech emotion recognition using support vector machine", electronic and mechanical engineering and information technology, international conference on Aug 12 -14, vol.2, pp.621-625, 2011.
34. Tsuta, M. Marsy, G. E, Sugiyama, T. Fujita, and K. Sugiyama, J, "Comaparison between linear discrimination analysis and support vector machine for detecting pesticide on spinach leaf by hyperspectral imaging with excitation-emission matrix", European symposium on artificial neural network – advance in computational intelligence and learning, Bruges (Belgium) on 22-24 April, pp.337-342, 2009.
35. Xia, B. Sun, H. and Lu, B. L, "Multi-view gender classification based on local gabor binary mapping pattern and support vector machines". IEEE International joint conference on neural networks. pp.3388-3395, 2008.
36. Vapnik, V, "Nature of statistical learning theory", New York, springer-verlag, 1995.
37. Zhang, J. Marszalek, M. Lazebnik, S. Schmid, C, "Local features and kernels for classification of texture and object categories: A comprehensive study", IJCV, Vol.73, pp.213–238, 2007.
38. Zhang, Y and Wu, L, "Classification of fruits using computer vision and a multiclass support vector machine", sensors, Vol. 12, pp.12489-12505, 2012.
39. Zadeh, L. A, "Fuzzy Sets", information and control, Vol.8, pp. 338-353, 1965.

Authors Biographies



R. Madhanagopal received B.Sc (Statistics), M.Sc (Statistics), M.Phil (Statistics) and M.B.A degree from Bharathiyar University, Coimbatore in 2002, 2004, 2006 and 2010. Currently, he is an Assistant professor and since 2010, Ph.D research scholar in Madras Christian College, Chennai, Tamil Nadu, India. His research interests include Multivariate Analysis, Data Mining, Fuzzy Logic and Data Envelopment Analysis.



R. C. Avinaash received B.Sc (statistics) degree from Madras Christian College, Chennai, Tamil Nadu, India in 2012. Currently doing M.Sc (Statistics) in Madras Christian College, Chennai, Tamil Nadu, India. His research interests include Multivariate Analysis, Data Mining.



K. Karthick received B.Sc (statistics) degree from Madras Christian College, Chennai, Tamil Nadu, India in 2012. Currently doing M.Sc (Statistics) in Madras Christian College, Chennai, Tamil Nadu, India. His research interests include Multivariate Analysis, Data Mining, Time series Analysis.