Available online at: https://ijact.in

Date of Submission	23/01/2019
Date of Acceptance	24/02/2019
Date of Publication	28/02/2019
Page numbers	3069-3074 (6 Pages)

ISSN:2320-0790

An International Journal of Advanced Computer Technology

PERFORMANCE ANALYSIS OF SVM WITH QUADRATIC KERNEL AND LOGISTIC REGRESSION IN CLASSIFICATION OF WILD ANIMALS

Suhas M.V[1], Swathi B.P[2]
[1]Department of ECE, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal 576104
[2]Department of I&CT, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal 576104
*swathi.bp@manipal.edu

Abstract: In an attempt to develop a system to classify the wild animals using image processing and classification techniques, we study the usage of Haralick textural features are used in wild animal classification which is a computer aided pattern recognition system. The Haralick features from two wild animal classes that include leopard and wildcat are extracted to from the image database. Support Vector Machine (SVM) with quadratic kernel function model and Logistic Regression (LR) model are developed and tested using the created dataset. In each case, the performance of the classifier is measured.We also compare the performances of SVM and LR with and without pre-processing the dataset using Principal Component Analysis (PCA). This study reveals an increment in the accuracy post pre-processing of the dataset.

Keywords: Haralick features, Support Vector Machine, Quadratic Kernel Function, Logistic Regression, Principal Component Analysis, Wildlife.

I. INTRODUCTION

The wildlife populations are increasingly endangered as human activities are manipulating natural systems through aggressive deforestation. Furthermore, the increasing urbanization has led to an increased entry of wildlife into human inhabitation. This has proven to be problematic situation for wildlife, human inhabitation and our natural systems. The reduction in the wildlife, their movement in the human inhabitation needs to be monitored through the help of technology

There has been development of various technologies by engineers during the past decades, for the monitoring of individual animals. These technologies include radio tracking [1], Global Positioning System (GPS) tracking [2][3][4],animal-mounted video monitoring systems [5], satellite tracking [6], and wireless sensor networks [7][8][9]. However, these efforts have been mostly carried out over a short period of time (often for a period of few hours, days or weeks),on a relatively small number of wildlife species, and over small geographical areas. Furthermore, data collected from sensor and camera by different individual and researchers are dispersed in time and space, isolated from each other and represented in various forms. With hardware and embedded computing technologic advances, existing visual monitoring technologies for wildlife monitoring using camera have developed to an extent where they are available commercially at a reasonable cost that are rapidly deployable, easy to maintain, and therefore used practically by a large number of non-professional citizens. These system in the background require a robust computer aided classification technique for effectively and accurately classifying the wild animals especially ones which look alike.

In an attempt to develop a robust computer aided wildlife tracking technique we test a system which will be able to classify two wild animal classes that includes leopard and wildcat. The dataset consists of pre-captured images of these two classes from 101 Caltech image dataset. The data points extracted from the images for this work belong to high-dimensional space. This is because the image is represented by its grey level values. Hence, we need a pattern recognition technique which adapts for a large dataset with multi dimension. Among the applications such as character recognition, image or signal assessment, Support vector machines are effectively used [10]. In this work, we compare the performances of Support vector machine (SVM) and logistic regression (LR) with and without pre-processing the dataset using PCA to classify the aforementioned classes.

II. BACKGROUND THEORY

This section furnishes the ground preparation required to implement the methodology. Entire section is divided among the theoretical aspects of techniques such as textural features, classification models such as Support Vector Machines and logistic regression and pre-processing technique i.e. Principal Component Analysis.

2.1. Textural features

Spectral, Textural and Contextual are the three features using which images can be interpreted [11]. In monochrome photographs, it is important to take into account textures and tones particularly in the context of independent processing of small image areas. In our work, the textural features are drawn out because they possess spatial distribution information entrenched as variations in tone within a band. The tone in an image is described as the change in grey shades of resolution cells which are considered as fine, smooth, irregular, course, lineated, and so forth. At the same time texture represent the statistical or spatial distribution of grey tones in an image and describes the lightness or darkness of a specific area of an image. Texture and tone have deeply complex relationship among each other [11]. Spatial-dependence matrix of grey tone represents the frequency of occurrence of one gray tone to another gray tone in a specified spatial relationship on an image. Textural features can appear as discrete histograms, scalar numbers or empirical distributions. The dominance of texture boosts with increment in number of discrete gray tone unique features of within the small area.

2.2. Support vector machine with quadratic kernel function and Logistic regression

SVM is a very powerful classifier which has been used in various applications [12]. SVM is a linear classifier which shows a exceptional performance during classification and regression in comparison to other techniques in machine learning[12]. It maximises the margin between two disjoint half spaces. SVM can be made to behave as a nonlinear classier by projecting the feature space into a high-dimensional feature space, where it constructs an optimal discriminant hyper plane using a nonlinear kernel function.
The model maps input data S={X}into a high-dimensional (possibly infinite dimensional) feature space F ={f (X)} by selecting an adequate mapping function f, a kernel. The performance of the SVM classifier significantly depends upon the selection of kernel function and it is purely application depended. The goal of the SVM is to decrease generalization error upper bound by increasing the hyper-plane and the data margin [13]. There are several types of kernels among which Radial Basis (RBF) and polynomial kernel functions are very popular. RBF is used when there is no prior knowledge about the dataset which is defined as a real-valued function whose value depends only on the distance from the origin, so that $f(x) = f(\|X\|)$. In absence of expert knowledge about data and domain, the Gaussian RBF kernel makes a good default kernel. This is because it subsumes polynomial and linear kernel. Linear Kernels and Polynomial Kernels are a special case of Gaussian RBF kernel. Polynomial kernel functions are used to map hyperbolic surface to a plane which is the reason for using such a kernel in our work. The value of sigma in Gaussian kernels, defines how the kernel will manage the neighbour and points that are not in its close vicinity. A high value of sigma takes into account all points but does not give priority to neighbourhood [14]. Medium Gaussian kernel takes benefit of both low and high values of sigma [15].
Logistic regression is a statistical method to infer the outcome variable [16]. To build a LR model, dataset should have one or more independent variables that determine an outcome. The outcome is measured with a dichotomous variable. LR is effectively applied in many applications such as verb expression's sentiment analysis, of physical activity enjoyment on motor ability, modelling crash likelihood information under rainy weather conditions, prediction of lung nodule biopsy method etc. LR is applicable when the when there is no multi co-linearity among the predictor variables and when there is no outlier in the dataset

2.3. Principal component Analysis

Principal component Analysis (PCA) is a data reduction technique through data transformation. PCA is unsupervised data reduction technique for multivariate data [17].PCA coverts the original dataset into another dimension in such a way that the original information is not lost. Unlike feature subset selection methods, PCA does not remove any attribute but transforms the features into dimensions such that the axes along which there is least variation, those features put together gives less information about the dataset.
If there is a feature vector X containing p number of features and n instances, after applying PCA, we obtain q number of features where q is much smaller than p. PCA does this by finding m linear combinations of principle components,pc1, pc2…….pcm such that pc1> pc2>…….>pcm. This means that variation along with pc1

is much greater than variation along pcm and it is up to the developer of the model to choose the number of principal components which in turn depends upon the performance requirement. The direction along the maximum variation is represented by the Eigen vector. The steps followed by PCA are: mean normalization to pre-process the dataset which helps to bring data points around the mean, computation of covariance matrix from normalised dataset, computation of Eigen vector from which we obtain the variation along several dimensions.

III. METHODOLOGY

The block diagram shown in the Figure1briefly illustrates the methodology of the work. The work that is proposed is conducted in a workstation with specifications: Intel® Core™ i5-6200U CPU 4GB RAM and 2.30GHz of clock frequency and 64-bit operating system. We have used MATLAB® version R2017a for implementation of the work.

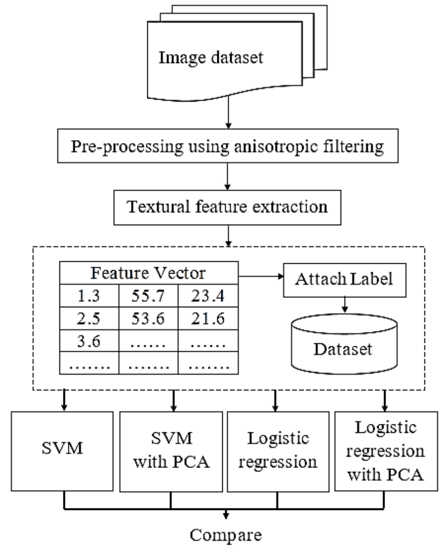


Fig. 1: Flowchart of Methodology

3.1. Dataset

The image data were obtained from the Caltech-101 dataset [18]. The two classes of object intended to be classified were leopard and wild cat. The database included 198 images of leopards and 34 images of wildcats. Figure 2 shows the example images obtained from the image database.

Fig. 2: Example images of leopards (top row) and wildcats (bottom row) used for the dataset obtained from Caltech-101 dataset.

3.2. Pre-processing

In the method, the classified images undergo pre-processing in order to clear away noise that might reduce the quality of features to be extracted.

The pre-processing step constitutes filtering images in order to clear away the implicit noise. Application of Gaussian filter expounds the smaller image features, hence leading to results which are false negative. Due to these threats, non-linear anisotropic filter is decided to apply. The applied anisotropic filtering preserves and smooth out the region that is homogeneous and the discontinuity of the inhomogeneous region are enhanced. Mathematical definition of Anisotropic filter [19] is as follows:

$$I_t = \mathrm{div}(c(x,y,t)\nabla I) = c(x,y,t)\Delta I + \nabla c.\nabla I \qquad (1)$$

Where div: divergence operator,∇: gradient operator and Δ is the Laplacian operator. The diffusion function of c(x,y,t) depends on the image intensity gradient magnitude unlike linear filter where the diffusion coefficient is considered constant independent of space location. The diffusion function used in our method is defined as:

$$c(\nabla I) = e^{(-(|\nabla I|/K)^2)} \qquad (2)$$

The diffusion function is responsible for improving the edges of high contrast over low contrast ones. The decision of selection of such a diffusion functions is dependent on its capacity to improve the high contrast edges. The constant K in equation (2) is set manually.

3.3. Feature extraction

In this work, a Gray-Level Co-occurrence Matrix (GLCM) is employed for extraction of features that figures the frequency of various combinations of pixel grey levels in an image. When a new image is given as input to the model, whose class is unknown the model assigns it to one of the known class. The models learnt for this work have been trained with 36 various textural features that are obtained from the greytone spatial dependence matrix [11]. The mathematical definitions of these feature metrics are defined as follows.

Energy:

$$f_1 = \sum_i \sum_j p(i,j)^2 \quad (3)$$

Contrast:

$$f_2 = \sum_{n=0}^{N_g-1} n^2 \left\{ \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j) \,\middle|\, |i-j| = n \right\} \quad (4)$$

Correlation:

$$f_3 = \frac{\sum_i \sum_j (i,j) p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y} \quad (5)$$

Sum of Squares (Variance):

$$f_4 = \sum_i \sum_j (i-\mu)^2 p(i,j) \quad (6)$$

Homogeneity:

$$f_5 = \sum_i \sum_j \frac{1}{1+(i+j)^2} p(i,j) \quad (7)$$

Sum Average:

$$f_6 = \sum_{i=2}^{2N_g} i p_{x+y}(i) \quad (8)$$

Sum Variance:

$$f_7 = \sum_{i=2}^{2N_g} (i-f_8)^2 p_{x+y}(i) \quad (9)$$

Sum Entropy:

$$f_8 = \sum_{i=2}^{2N_g} p_{x+y}(i) \log\{p_{x+y}(i)\} \quad (10)$$

Entropy:

$$f_9 = \sum_i \sum_j p(i,j) \log(p(i,j)) \quad (11)$$

Difference Variance:

$$f_{10} = varience\ of\ P_{x-y} \quad (12)$$

Difference Entropy:

$$f_{11} = -\sum_{i=0}^{N_g-1} p_{x-y}(i) \log\{p_{x-y}(i)\} \quad (13)$$

Autocorrelation:

$$f_{12} = \sum_i \sum_j (i,j) p(i,j) \quad (14)$$

Dissimilarity:

$$f_{13} = \sum_i \sum_j |i-j| p(i,j) \quad (15)$$

Cluster shade:

$$f_{14} = \sum_i \sum_j (i+j-\mu_x-\mu_y)^3 p(i,j) \quad (16)$$

Cluster prominence:

$$f_{15} = \sum_i \sum_j (i+j-\mu_x-\mu_y)^4 p(i,j) \quad (17)$$

Maximum probability:

$$f_{16} = \max_{i,j} p(i,j) \quad (18)$$

Information Measure of Correlation:

$$f_{17} = \frac{HXY - HXY1}{\max\{HX, HY\}} \quad (19)$$

$$f_{18} = (1 - \exp[-2(HXY2 - HXY)])^{1/2} \quad (20)$$

$$HXY = -\sum_i \sum_j p(i,j) \log(p(i,j)) \quad (21)$$

Where HX and HY are entropies of P_x and P_y

$$HXY1 = -\sum_i \sum_j p(i,j) \log\{p_x(i) p_y(j)\} \quad (22)$$

$$HXY2 = -\sum_i \sum_j p_x(\mathrm{i}) p_y(\mathrm{j}) \log\{p_x(i) p_y(j)\} \quad (23)$$

For a given distance d a set of four angular gray-tone spatial-dependency matrices are found, which obtains a set of four values for each 18 measures. The classifier inputs consists of a set of 36 features that are obtained from the mean and range of 18 measures, averaged over the four values. The extracted 36 features were inspected to observe that there are no outliers, missing values and noise; hence a clean dataset is obtained.

IV. RESULT ANALYSIS

In the following section we analyze the result obtained in our experiments. The performance of the system is studied by finding the performance parameters such as accuracy, precision and recall. A plot of receiver operating characteristic curve (ROC)is also shown for understanding the performance.

4.1. Confusion Matrix and ROC

The confusion matrices obtained for SVM, SVM with PCA, logistic regression and logistic regression with PCA are shown in the tables 1, 2, 3 and 4.

Table 1: Confusion Matrix for SVM with quadratic kernel

		Predicted class	
		Leopard	wildcat
True class	Leopard	197	1
	wildcat	9	25

Table 2: Confusion Matrix for SVM with quadratic kernel and PCA

		Predicted class	
		Leopard	wildcat
True class	Leopard	198	0
	wildcat	8	26

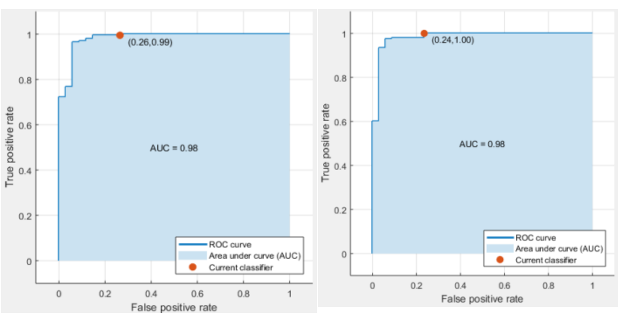
Table 3: Confusion Matrix of Logistic regression accuracy

		Predicted class	
		Leopard	wildcat
True class	Leopard	196	2
	wildcat	11	23

Table 4: Confusion Matrix for Logistic regression with PCA

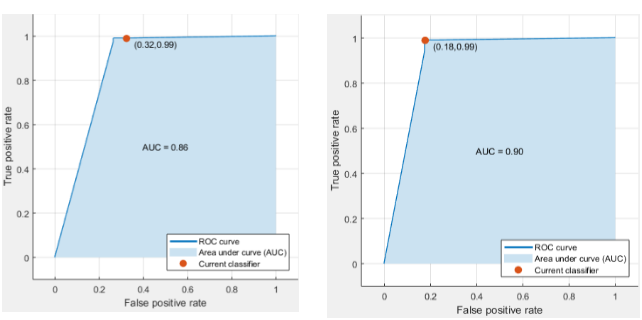
		Predicted class	
		Leopard	wildcat
True class	Leopard	196	2
	wildcat	6	28

5 fold cross-validation technique is used to measure the accuracy of the model. 5 folds are chosen because the dataset used for experimentation is comparatively smaller in size. In m-fold validation technique, the entire data set is divided into m folds. Every time a training model is built, m-1 folds are used to train the model and one fold is used for testing purpose. The entire process is carried out m times such that the subset used in training is not used in testing the model. Figure 3 and Figure 4shows the obtained ROC curve for SVM and Logistic Regression respectively.



(a) (b)

Fig. 3: ROC plot for SVM (a) without PCA(b) with PCA: 20/36 feature selected.



(a) (b)

Fig. 4: ROC plots for Logistic regression(a) without PCA (b) with PCA: 20/36 features selected.

4.2. Accuracy Precision and Recall

To understand the quantitative performance of the system, we calculated the accuracy, precision and recall from the data tabulated on the confusion matrix as per the following procedure. Accuracy is measured using the formula$Accuracy = \frac{TP+TN}{Total}$. Precision is measured using the formula$Precision = \frac{TP}{TP+FP}$. Similarly the sensitivity or recall was also measured using the formula $sensitivity = \frac{TP}{TP+FN}$.The observations are tabulating in the table 5.

Table 5: Tabulation of Precision and recall

	Accuracy	Precision	Recall
SVM	95.7%	96.15%	73.5%
SVM with PCA	96.6%	100%	76.4%
Logistic Regression	94.4%	92%	67.6%
Logistic regression with PCA	96.6%	93.3%	82.3%

4.3. Prediction speed and training time

The prediction speed based on number of observations per second and the training time is tabulated in table 6. These parameters though are variable from iteration to iteration the average lies within the values shown here.

Table 6: Prediction speed and training time

	Prediction speed	Training time
SVM	6800 obs/sec	0.60186 sec
SVM with PCA	2300 obs/sec	0.55620 sec
Logistic Regression	4800 obs/sec	1.1273 sec
Logistic regression with PCA	1800 obs/sec	1.5084sec

V. CONCLUSION

The best performance is achieved with SVM after pre-processing the dataset using PCA with an accuracy value 96.6%, precision of 100%, sensitivity of 76.4% and ROC value 0.98. The performance of Logistic regression with PCA though is close to SVM with PCA, in terms of accuracy, the area under the curve reveal that SVM with PCA outperforms logistic regression with PCA.

This work which is focused on developing classification models for wild animals through its results proves useful in monitoring the different classes of animals. The work can be extended by including many such wild animals which needs to be monitored for many reasons using the camera tracking systems.

VI. REFERENCES

[1] L. D. Mech, A Handbook Of Animal Radio-tracking, Univ. Of Minn. Press, Oxford, 1983.

[2] P. Juang, H. Oki, Y.Wang, M. Martonosi, L.-S. Peh, and D. Rubenstein, “Energy-efficient computing for wildlife tracking: design tradeoffs and early experiences with zebranet,” in Tenth International Conference on Architectural Support for

Programming Languages and Operating Systems (ASPLOSX),San Jose, CA, 2002.
[3] Jon Young and Tiffany Morgan, Animal Tracking Basics, Stackpole Books, 2007.
[4] Ian A.R. Hulbert and John French, “The accuracy of gps for wildlife telemetry and habitat mapping,” Journal of Applied Ecology, vol. 38, pp. 869–878, August 2001.
[5] R. J. Moll, J. Millspaugh, J. Beringer, J. Sartwell, and ZhihaiHe, “Animal-borne video systems: a new era of behavioralecology,” Trends in Ecology and Evolution, vol. 22, pp. 660–668, November 2007.
[6] M. Williams, A. Lunsford, D. Ellis, J. Robinson, P. Coronado, and W. Campbell, “Satellite tracking of threatened species,” Argos Newsletter, vol. 53, 1998.
[7] I.F. Akyildiz, Y. Sankarasubramaniam W. Su, and E. Cayirci,“Wireless sensor networks: a survey,” Computer Networks, vol. 38, pp. 393–422, March 2002.
[8] H. Gharavi and K. Ban, “Vision-based ad-hoc sensor networksfor tactical operations,” in World Wireless Congress, 3G Wireless2002, San Francisco, 2002.
[9] Robert Szewczyk, Alan Mainwaring, Joseph Polastre, and David Culler, “An analysis of a large scale habitat monitoring application,” in In Proceedings of the Second ACM Conference on Embedded Networked Sensor Systems (SenSys), 2004.
[10] Chen, Szu-Ta, Kathiravan Srinivasan, Chen Lin, and Chuan-Yu Chang. "Neonatal Cry Analysis and Categorization System Via Directed Acyclic Graph Support Vector Machine." In Big Data Analytics for Sensor-Network Collected Intelligence, pp. 205-222. 2017.
[11] Robert M. Haralick, K. Shanmugan, and ItshakDinstein, Texture features for image classification, in IEEE Transactions on Systems,Man, and Cybernetics, Vol.Smc-3, No.6, pp.610 - 621, (1973)
[12] Meyer, David, Friedrich Leisch, and Kurt Hornik. "The support vector machine under test." Neuro computing 55, no. 1, 169 - 186, (2003)
[13] Amari, Shun-ichi, and Si Wu. Improving support vector machine classifiers by modifying kernel functions." Neural Networks 12, no. 6, 783 - 789, (1999)
[14] Blanc-Talon, Jacques, Don Bone, Wilfried Philips, Dan Popescu, and Paul Scheunders, eds. Advanced Concepts for Intelligent Vision Systems: 12th International Conference, ACIVS 2010, Sydney, Australia, December 13 - 16, 2010, Proceedings.Vol. 6474. Springer, (2010).
[15] Dagher, Issam. "Quadratic kernel-free non-linear support vector machine." Journal of Global Optimization 41, no. 1: 15-30. 2008
[16] HosmerJr, David W., Stanley Lemeshow, and Rodney X. Sturdivant. Applied logistic regression. Vol. 398. John Wiley & Sons, 2013.
[17] Jolliffe, Ian. "Principal component analysis." In International encyclopedia of statistical science, pp. 1094-1096. Springer, Berlin, Heidelberg, 2011.
[18] L. Fei-Fei, R. Fergus and P. Perona. Learning generative visual models from few training examples: an incremental Bayesian approach tested on101 object categories. IEEE. CVPR 2004, Workshop on Generative-Model Based Vision. 2004
[19] P. Perona and J. Malik, Scale-space and edge detection using anisotropic diffusion, IEEE Transactions on Pattern Analysis and Machine Intelligence, 12(7), 629 - 639,(1990)