International Journal of Scientific & Engineering Research, Volume 7, Issue 3, March-2016
ISSN 2229-5518

Evaluating Kernel Effect on Performance of SVM Classification using Satellite Images

Anand N.Khobragade, Mukesh M. Raghuwanshi, Latesh Malik

Abstract— Support Vector Machine (SVM), a statistical learning algorithm has proved its excellencies in almost every domain towards of Image classification on various data formats. However, satellite images are more complex due to its number of bands that really contribute to image classification problem. Supervised approaches for satellite images needs more precise, fast and efficient machine learning models. Even though, SVM is non-parametric binary classifier, but could be mould to resolve multi-class classification problems. SVM comprise of superior machine learning algorithms group based upon the mean field theory, which basically uses quadratic optimization technique. SVM kernels are used to demarcate optimum hyperplane in non-linear problem space, which in return generalize to unseen samples with least error among all possible boundaries separating two classes. Kernel efficiency varies with SVM implementations like C-SVC, nu-SVC, nu-SVR, single class SVM, and CS-SVM. Use of density estimation function along with various norms, have great impact on SVM kernel performance for classification problems. Kernel tricks very well handle the problem of mixed pixels, which is vital affecting source on classification accuracies. Research illustrate that SVM kernel type sigmoid perform better than other SVM kernel types like Polynomial, Radial Basic, and Linear kernel at fixed training data size of 300 pixel/class. It is most appropriate in situations where the training sample data are difficult to collect, as it works better even with small number of training samples. Perhaps, 75 to 100 pixels per class are recommended for accuracy assessment of classification.

Index Terms—Support vector machine, Kernel efficiency, Satellite image, Multispectral, Feature extraction, nu-SVC, Image classification, CS-SVM, Multi-class classification problem

1 Introduction

AGRICULTURE is backbone of Indian economy as majority of population lives in rural areas. Traditionally, agricultural production estmatations are based upon its sown area, which is difficult to manage using available mechanism. The prominent causes for inaccurate and opaque figures on Indian agronomy are existing inadequate facilities, unstable mechanism, and sluggish functionaries [1]. Using remote sensing technologies, researchers are hopeful towards finding solutions of such problems. It would be great challenge to classify satellite images that comprise of multi-spectral bands,which are more complex to understand, process, and classify in return[22].

Multispectral images are acquired by means of remote sensing radiometers or sensors. Dividing the spectrum into many bands, multispectral is the opposite of panchromatic, which records only the total intensity of radiation falling on each pixel. Usually, satellite images captured in three or more channels. Each one used for acquiring single digital image in a small band of visible spectra, ranging from 0.7 µm to 0.4 µm, called red-green-blue (RGB) region, and going to infrared wavelengths of 0.7 µm to 10 or more µm. Remote Sensing Images are considered as most complex in nature as regards to image classification.

- *Anand N Khobragade, Research Scholar, GH Raisoni College of Engineering and Research, Digdoh Hills, Nagpur, India. PH-07122220086 E-mail: semanand@gmail.com*
- *Mukesh M Raghuwanshi, Professor, Computer Technology Department, YC College of Engineering & Research, Wanadonagri,Nagpur,India*
- *Latesh Malik, Head of Department, Computer Science & Engg, GH Rasoni College of Engineering and Research, Digdoh Hills, Nagpur, India.*

Purely statisitcal based algorithms suffer from poor support for high dimensionality. Being very slow in computations, fuzzy or neural based learning algorithms not feasible for big data computing, whereas artificial neural network based learning algorithms are ill with generalization problems. However, the most realistic solution towards all above cited lacunes is use of hybrid approach, which will encapsulate strengths of not only statistical but learning algorithms [2]. Statistically Learning Algorithm surmount these drawbacks with its capabilities like high computational efficiency, robust in high dimensionality, good in generalization and hence works well with small training dataset, controls accuracy vs complexity in function estimation, and superior performance in classifying hyperspectral images.

The statistical learning theory provides a framework for studying the problem of gaining knowledge, making predictions, making decisions from a set of data. In simple terms, it enables the choosing of the hyper plane space such a way that it closely represents the underlying function in the target space. In statistical modeling we would choose a model from the hyspothesis space, which is closest (with respect to some error measure) to the underlying function in the target space. More on statistical learning theory can be found on introduction to statistical learning theory [28].

Support vector machines (SVM) are nowadays very popular amongst researchers for addressing spectrum of remote sensing applications. However, SVM is a binary, parametric, supervised, statistical learning classifier and very sensitive to the parameters setting as well as choice of training sites. Self-training is an effective semisupervised method, which can reduce the effort needed to prepare the training set by training

INTERNATIONAL JOURNAL OF SCIENTIFIC & ENGINEERING RESEARCHVOLUME 6, ISSUE 12, DECEMBERS-2015
ISSN 2229-5518

the model with a minimum number of labeled traning pixels and an additional set of unlabeled examples. A novel semisupervised SVM model was developed which uses self-training approach for addressing the problem of remote sensing land cover classification [3].

It is studied that nearly all researches are based on binary classification and focus on how to estimate the generalization performance of SVMs effectively and efficiently. For problems with more than two classes, where a classifier is typically constructed by combining several binary SVMs, most researchers simply select all binary SVM models simultaneously in one hyper-parameter space. The studies reveal that all-in-one method works well, as compare to another one-in-one method where each binary SVM model is selected independently and separately[23][24]. It is studied that, pair-wise coupling (PLC) multi-class approaches works in two steps: First the original pairwise probabilities are converted into a new set of pairwise probabilities, then pairwise coupling is employed to construct the global posterior probabilities. Experimental results show that this algorithm is effective and efficient.[25][26].It is learnt that fuzzy based multi-class approach is also adapted for converting Binary SVM problem into multi-class problem[27].

A kernel is an integral part of SVM, which actually drives the accuracy of image classification. Usually, standard model selection methods are applied for choosing a kernel such as cross validation of various kernel functions [7].

The study divulges the vital aspects of kernels affecting performance and its applicability in agriculture domain towards classification of remotely sensed data. This paper is organized as follows. Section II present an overview of Support vector machine(SVM). Kernel effects are discussed in Section III, whereas means & methods were demonstrated in Section IV followed by evaluation methodology and conclusions in subsequent sections.

2 Support Vector Machine

Classification of remote sensing images is considered as a complex task as even minute feature extraction also needs to be taken in consideration. More precise machine learning algorithms required to be applied on supervised classification of remote sensing images which ensure to give fast and efficient results. SVM has given its satisfactory results to researchers all over the world as far as Remote Sensing Images are concern.Support Vector Machine (SVM) is basically a classifier function that does classification by constructing hyperplanes in a multidimensional space which separates cases of different class labels[4, 12].

SVM accuracy depends upon the major factors like size of training sets, number of features used, and type of kernel, which plays vital role in satellite image classification[22]

In statistical learning theory the problem of supervised learning is formulated as follows. We are given a set of training data {(x1,y1)... (xl,yl)} in Rn × R sampled according to unknown probability distribution P(x,y), and a loss function V(y,f(x)) that measures the error, for a given x, f(x) is "predicted" instead of the actual value y. The problem consists in finding a function f that minimizes the expectation of the error on new data that is, finding a function f that minimizes the expected error: $\int V(y, f(\mathbf{x}))\, P(\mathbf{x}, y)\, d\mathbf{x}\, dy$ [28]

2.1 Learning and Generalization

Early machine learning algorithms aimed to learn representations of simple functions. Hence, the goal of learning was to output a hypothesis that performed the correct classification of the training data and learning algorithms were designed to find such an accurate fit to the data [29]. The ability of a hypothesis to correctly classify data not in the training set is known as its generalization. SVM performs better in term of not over generalization when the neural networks might end up over generalizing easily [30]. It is needed to see a situation to make the best trade-off in trading complexity with the number of epochs; the following illustration highlights the facts;

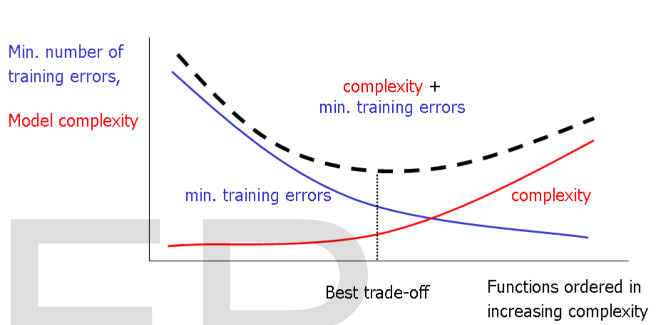

Fig.1. Number of Epochs Vs Complexity

If standard quadratic problem solver is used for SVM training then it automatically involves solving a huge QP problem even though the data size is bare minimum. The computation of m × m matrix in memory, helps to shorten the size of problems to which SVM could be applied efficiently.

3 Linearity & kernel effect

Usually, a separating hyperplane is used when data is linearly seperable. However, its not always true, as real world problems are almost complex and non-linear in nature. Hence, machine learning approaches used Kernel trick to map the non-linear input data to higher dimensional space so as to make it simpler to separate linearly [30]. Support vector machines uses an implicit mapping Φ of data as input into high-dimensional feature space as defined by kernel functions, which returns the inner two data points in the feature space. Such mapping is defined by Kernel function and its better illustrated in following figure;

$$(K(x,y)) = \mathbf{\Phi}(x).\,\mathbf{\Phi}(y)$$

IJSER

INTERNATIONAL JOURNAL OF SCIENTIFIC & ENGINEERING RESEARCHVOLUME 6, ISSUE 12, DECEMBERS-2015
ISSN 2229-5518

Transforming the data into feature space makes it possible to define a similarity measure on the basis of the dot product. As the appropriate features space more relevant will be pattern recognition from the image.

$$\langle x_1, x_2 \rangle \longleftarrow (\mathrm{K}(x1, x2)) = \Phi(x1) . \Phi(x2)$$

Estimating parameters w & b will resolves simple linear scenario in which data is separated by a hyper plane. The Kernel trick allows SVM's to form nonlinear boundaries. The dot product of nonlinearly mapped data can be expensive. In some cases, there is tremendous explosion of dimensionality of feature space.

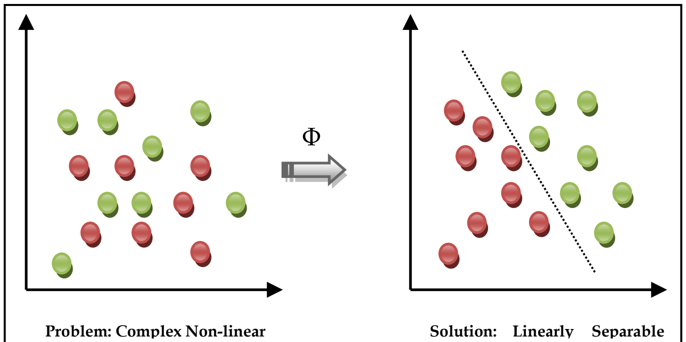

Fig.2. Kernels's mapping trick

For example, if we aims to construct decision surface corresponding to polynomial of degree 2, then it gives rise to n(n+3)/2 coordinates in feature space, whereas if degree of polynomial exceeded to 5, then coordinates in feature space may toll upto billions. Hence, kernel functions are then converted into dual form for simplified optimization. The kernel trick just picks a suitable function that corresponds to dot product of some nonlinear mapping instead [29]. Irrespective of type of kernel to be mapped, use of specific kernel function is feasible for any dimensional space without addition in cost of computation.Kernel methods also address the problem of data loss during the process of feature extraction.

3.1 Kernel Function

The idea of the kernel function is to enable operations to be performed in the input space rather than the potentially high dimensional feature space. Hence the inner product does not need to be evaluated in the feature space. It is needed a function to perform mapping of the attributes of the input space to the feature space. The kernel function plays a critical role in SVM and its performance. It is based upon reproducing Kernel Hilbert Spaces [29].

$(\mathrm{K}(x1, x2)) = \Phi(x1) . \Phi(x2)$, where

K is a symmetric positive definite function, which satisfies Mercer's Conditions. The kernel represents a legitimate inner product in feature space. The training set is not linearly separable in an input space, but separable in the feature space.

$$K(x, x') = \sum_{m}^{\infty} a_m \phi_m(x) \phi_m(x'), \quad a_m \geq 0,$$

$$\iint K(x, x') g(x) g(x') dx dx' > 0, \quad g \in L_2$$

Most of the kernel functions are based upon the convex optimization. Convex optimization uses minimization techniques which really makes the optimization problem much easier than general one. The strong assumption is made that any local minima must have global minima in convex functions[8]. The studies show that it is very difficult to select any one kernel which gives exact generalization (Vapnik, 1995). Researchers studied different kernels for investigating the effect of right choice of kernel on accuracy of classification using multispectral data and found that the radial basis as well as the linear splines performed equally well and acquired highest accuracy for their data set, et al.(Pal, 2002). It was proposed to use new kernels, especially suitable for spatial and spectral characteristics of remote sensing data towards land cover classification and found it more efficient than RBF kernel [20].

3.2 Choice of Kernel

Choosing a kernel function for SVM classification plays a crucial role, especially when input data is noisy. The kernel function to be chosen should have two properties as it must caught the measure of similarity appropriate to specific domain and it should be evaluated in less computation than required in an explicit evaluation of the corresponding feature mapping [18].

The linear kernel, polynomial kernel, Radial Basis Function (RBF) kernel, linear spline, Fourier, splines, B-splines, additive kernels and sigmoid kernel are major kernels in context with SVM implementation.

The linear kernel is formulated by transposing input vector and then dot product with original input. The polynomial kernel will use degree to control computation in feature space. It represents the similarity of transformed training vectors in a feature space over polynomials of the original input vectors, allowing learning of non-linear models. RBF is popularly known as Guassian radial basis kernel, which uses similarity measures for ready interpretations. The feature space of the RBF kernel may have infinite number of dimensions, which control by its σ parameter. It also produces a piecewise linear solution which can be attractive when discontinuities are acceptable [29] and hence could be the best for agriculture crop classification problems. Fishers kernel is another similarity based kernel which estimates the unknown classes by bringing close to known classes. Graph kernel is basically based upon structure mining that computes inner products of vectors on graphs. String kernels operates on string as a function measuring the similarity of pairs of string[8]. Sigmoid kernel uses the function thresold to activate the input vector to transform the input to bounded range and calculates a positive derivative at each point. The formulation of kernels that most relevant to LIBSVM are as follows;

Linear Kernel: K(xi,xj)= xi^T.xj
Polynomial of degree p: K(xi,xj)= $(1+ xi^T.xj)^p$

IJSER

INTERNATIONAL JOURNAL OF SCIENTIFIC & ENGINEERING RESEARCHVOLUME 6, ISSUE 12, DECEMBERS-2015
ISSN 2229-5518

Radial Basis Function (RBF) kernel

$$K(\mathbf{x_i}, \mathbf{x_j}) = \exp(-\frac{\|\mathbf{x_i} - \mathbf{x_j}\|^2}{2\sigma^2})$$

Sigmoid Function: K(xi,xj)= tanh(β0xi^T.xj + β1)

4 Means and methods

Across the globe, researchers have applied variety of kernel methods on numerous data sets for wide applications ranges. Multiple Kernel Learning is not restricted by deterministic or statistical modelling assumptions and, therefore, is more flexible for modelling heterogeneity at different scales and integrating data and knowledge[24].

Research findings in [2] further illustrate that SVM sigmoid kernel perform better than other SVM kernel types like Polynomial, Radial Basic, and Linear kernel for non-variable training sample size of 300 pixel/class.

Multiple kernel ensembling is applied for feature extractionn using AVIRIS Hyperspectral data and other experiment done on fused image of DSM and LiDAR data so as to construct more effective training data pool that later be feeded to Active Learning Algorithms for interactive labeling. It is found that such intermingling is better than applying simple MKL approach [5]. Camps-Valls and L. Bruzzone [14] demonstrated graph based connectivity kernel and RBF kernel for optimizing primal of SVM and found to be much reliable, accurate than traditional approaches.

Sun & Li [18] proposes scalling of kernel function instead of estimating support vectors in feature space. The region around the input space is first get hold and then kernel function is scaled up correspondingly. The findings concluded with no neccessity of training for boundary conditions as seperating hyperplane. Zhi He & Junbao [23] evaluated the degree of agreement between Kernel and its classification by virtue of optimizing the combination of multiple kernels in merit of centered kernel alignment (CKA). In addition, they optimize the coefficients of data-dependent kernel (DK) using Fisher's discriminant analysis (FDA).

Demyanov & Christie [24] make use of multiple kernel learning approaches for calibrated reservoir modelling considering porous property distributions in sub-surface reservoir. MKL based history matching results were found better as compared to EnKF and kernel PCA with stochastic optimisation. E Ishikawa and others [26] demonstrate the use of local higher-order moment kernel (LHOM kernel) in SVMs for texture classification and when compare it with SVMs using other conventional kernels, it was obvious that LHOM achieve better trainability and give stable response to the texture classes with minimum support vectors and hence it could exhibits better class separability in the nonlinearly-mapped feature space. E Li, Daoliang, Zhao, and Chunjiang [25] use improved compound kernel function that has a higher accuracy of classification on Remote Sensing images. In addition, compound kernel improves the generalization and learning ability of the kernel.

Tuia, Matasci, and others[9] have proposed an wrapper method that intermingle feature selection and classification within the framework of multiple kernel learning (MKL). This embedded approach aims at capturing relevant features of satellite images for spatio-spectral automatic classification using data specific dedicated kernels for different features. Optimization of the optimal linear combination of kernels is done with the help of gradient descent on the SVM objective function. Bor-Chen Kuo, Hsin-Hua Ho[10] demonstrated the method for high resolution hyperspectral image classification using SVM with multiple kernels. This multi-kernel SVM classifiers then tested with the help of Reflective Optics System Imaging Spectrometer (ROSIS) data with 115 bands to evaluate the performance and accuracy of the proposed multi-kernel classifier.

Gurram and Kwon [17] developed a generalized Kernel-based Ensemble Learning (GKEL) algorithm for Hyperspectral image classification problems. This algorithm aims at generalizing the Sparse Kernel-based Ensemble Learning (SKEL) technique using optimally the sparsely weights and integrate an ensemble of targetted SVM classifiers which automatically conducted learning using randomly selected spectral feature sub-space with the use of a Gaussian kernel[16].

In 2014, Yadong Mu and Gang Hua [7] invented an algorithm which uses compact hash bits to enhance the efficiency of non-linear SVM kernel in very large scale visual classification problems. The samples containing compact hash bits were presented, upon which an inner product can be defined to serve as the surrogate of the original non linear kernels. Later, the nonlinear SVM problem have been resolved by transforming it into a linear SVM problem over the hash bits. A novel hashing scheme for arbitrary non-linear kernels is porposed as a critical component of Hash-SVM, via random subspace projection in reproducing kernel Hilbert space.

In addition, Bor-Chen Kuo, Hsin-Hua Ho[10] in 2014 put forward novel idea of a kernel-based feature selection method with a criterion that is an integration of an automatic method for selecting the radial basis function (RBF) parameter for SVM and the linear combination of features. This newly developed method achieves two properties that are the ranking of features according to the magnitude of coefficient and features with small subset to be calculated.

Further researchers continued the statistical analysis of different kernel methods namely Polynomial kernel, RadialBasis Function (RBF) kernel and Multilayerperceptron (MLP) kernel used for training SVM [12]. When multiple traits are coupled together at feature/ score/ decision level, it results in building accurate multimodal systems. Later, these findings be helpful for weighted match score to recognize an individual which actually improve the rate of recognizing an individual. The training time required for SVM using all three kernel methods is rooted in statistical analysis as well upon the performance curve in terms of parameters as recognition rates i.e. Genuine Acceptance Rate (GAR) and False Acceptance Rate (FAR) of SVM coupled system. They found that RBF kernel based SVM fused system gives better results as it requires the lowest training time as compared to other kernel methods. Also the recognition performance of RBF based SVM system is more than that of other kernel based systems [22].

INTERNATIONAL JOURNAL OF SCIENTIFIC & ENGINEERING RESEARCHVOLUME 6, ISSUE 12, DECEMBERS-2015
ISSN 2229-5518

5 Evaluation Strategy

Entire cycle of image classification is to be adopted for evaluating kernel effect on SVM performance. Initially, input vector is created from multi-spectral satellite images. Feature vector is applied and again it must be exported to .CSV file or any compatible format commensurate with desired software, may be even geotiff image sometimes. Input vectors are lables for known samples so as to train our SVM model. The important step is to make appropriate choice of kernel functions, which significantly varies with the nature of application or complexity of classification problem. Basic kernel effects is been evaluated using multi-spectral satellite images.

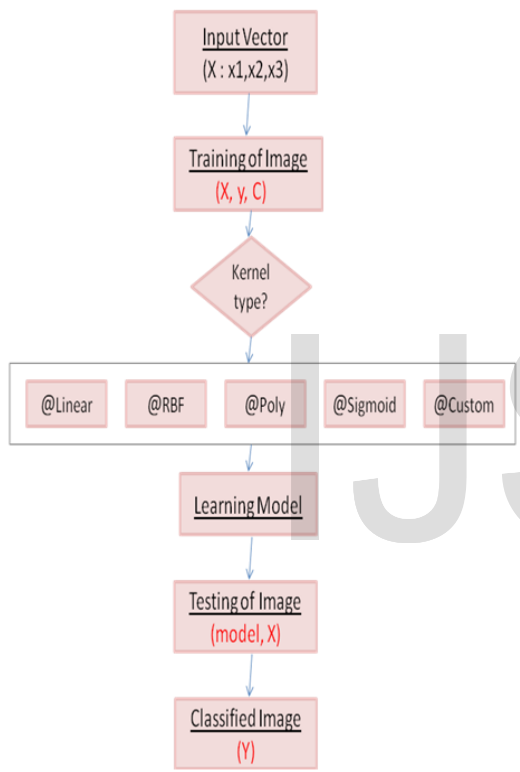

Fig.3. Evaluation Methodology for SVM Kernels

Classification accuracies of the remote sensing images will solely depends upon the parameter settings of the kernel chosen. The experimentation is performed upon the four kernels for classification problem. From the experimentation results, it is obvious that linear kernel is deemed to be more fitted into the training & testing if the multi-spectral image chosen for experiomentation. On countrary, time taken for classification is much more than other kernels. Rest three kernels viz. Sigmoid kernel, Intersection kernel, and Anova kernel performed as equal in terms of accuracy & time complexity, whereas Sigmoid and Anova kernel's performance is almost identical. This research exhibits the fact about the multi-spectral satellite dataset that used in this experimentation, is linearly seperble.

Kernel Type	Accuracy	Time Taken
LINEAR	92.36 %	141.618sec
SIGMOID	66.57 %	35.575sec
INTERSECTION	69.30 %	35.827sec
ANOVA	66.56 %	35.789 sec

In another experiment, we used our own java implementation for SVM classification, where the purpose was to check whether the satellite data was linearly seperable or complex to do so. C-SVC implementation of LIBSVM is used and tested the time complexity pertaining to traning & learning for crop classification remote sensing image. The experimentation parameter setting for SVM RBF kernel is mentioned below;

No. of pixels for Cotton Crop	: Varying with training sites
Standard deviation of samples	: 2
Number of iterations	: 2
Neighbourhood pixel pattern	: 8 neighbors
SVM kernel type	: RBF Kernel
Gamma Value for RBF Kernel	: 0.333
Penalty parameter	: 100

From results, it is evident that Linear kernel took 0.386 sec, whereas RBF kernel 0.846 sec for training SVM with minimum three basic features of remote sensing images. SVM model parameters are estimated as a outcome of training. Once started learning, the testing time for Linear kernel found to be 28sec and that of RBF is 1132sec.

IJSER

6. Conclusion

Studies reveals that SVM is the best classification model as compared to Neural Network, Fuzzy Logic, Decision Tree classifiers or any other machine learning algorithms. Inspite of SVM world wide accetance by researchers, non-linear kernels pays vital contribution to data depedent classification problems. Besides researchers have invented kernels for making variety of data formats seperable, scientist still have challenge to devise new kernels that may be appropriate for classification problems in particular. Simillarly, researchers reveals another variant of SVM kernel implementation that uses the estimation of probability density functions (e.g. Paolo, Gabriele and Sebastiano, 2005) with the help of easy and efficient machine learning procedures based on Mean Field theory. Density estimation based SVM when fused with fuzzy weighted matrices approach to define fuzzy kernels, produce better results. However, RBF kernel is found to be suitable for most the applications, except its overfitting problems in exceptional cases.

SVM kernel have been analyzed against most of the strategies for satellite image classification problems using range of remotely sensed images, viz. Multi-spectral ,Hyperspectral, or even Microwave Remote Sensing based satellite images. Most of the research based upon the multi-spectral satellite images, but study of hyperspectral remote sensing is emerging fast for vegetation analysis, especially agricultural crop mapping, acreage estimation, and yield productivity. The researcher further divulges that SVM demonstrate superior performance in classifying hyperspectral images acquired from AVIRIS, AVHRR, and MODIS sensors because of its suitability for high dimensional data.

With the advent of study of various SVM kernels functions, observation regarding classification of satellite image showed neither every kernel function is accurate over a specific data nor any specific kernel function will show its 100% accuracy over every data.Default choice of kernel is either Gaussian or polynomial kernel. If ineffective, more elaborate kernels are needed, where domain experts can assists in formulating appropriate similarity measures, and devise new kernel thereof. The appropriate choice of kernel parameters is crucial e.g. σ in Gaussian kernel, which is the distance between closest points with different classifications. In the absence of reliable criteria, applications rely on the use of a validation set or cross-validation to set such parameters.

Generalization of SVM could be ideally applied in event of small dimensionality of feature space, large separating margin, and small number of support vectors. SVM locates a separating hyperplane in the feature space and classify points in that space by means of iner dot products explicitely by kernels. It governs by a simple convex optimization problem which is guaranteed to converge to a single global solution, , where a relatively small number of mislabeled examples can dramatically decrease the performance.

SVM kernels has extensive adequacy in agriculture domain, especially when used in amalgamation with vegetation indices like NDVI, SVI, TaVI,etc. LIBSVM facilitates with provision of using Linear, Polynomial, RBF, and Sigmoid along with various norms viz. Euclidian norm, Diagonal norm, and Mahalonobis norm. Mahalonobis norm is suitable for Forest mapping, whereas Diagonal norm gives best results for Agricultural and Fallow Land mapping. While modeling the spatial contextual information for hard classifiers using Markov Random Field it has been found that Metropolis algorithm is easier to program and it performs better when compared with the Gibbs sampler [2]. Further, it has been found that in case of soft contextual classification Metropolis algorithm fails to sample from a random field efficiently and Gibbs Sampler performs better than the Metropolis algorithm, due to high dimensionality of the soft classification output. Further articulate that Metropolis algorithm is not suitable for contextual satellite image classification as it suffers from poor convergence & long computational time period. (Amitava Dutta, Anil Kumar and Soma Sarkar, Feb 2010)

With the experimentation done on multi-spectral remote sensing images, SVM kernels were tested for accuracies & classification time using MATLAB. It is disclose from the results that Linear kernel with accuracy of 92.36 %, performs well as compared to rest three kernels with the classification time as 141.618sec, which is the most costly than other kernels in completion. Rest of the kernels Intersection, Sigmoid and Anova atteained average accuracy of 67% with avg. classification time of 35 sec. Sigmoid and Anova kernel's performance is almost indistinguishable.

When performed experimentation with own SVM java implementation training & testing time using three RBG features of satellite images, the results were promising. Data found to be linearly seperable as Linear kernel was muct faster in both the phases of classification. Non-linear kernel may seem to be suffered from overfitting and hence delayed in SVM training and classification as well.

Use of Evolutionary Computing Techniques for feature dimension reductionality and optimizing SVM kernel parameter settings would be the future scope for continuing this research on SVM kernel. Evolutionary algorithms like Differential Evolution Algorithms (DE), Genetic Algorithm, Comprehensive Learning Particle Swarm Optimization (CLPSO) , may impact upon optimizing kernel functions for classification problems. The last but not the least, futuristic scope towards expanding horizon of this research would be the generation of crop acreage and crop yield model based upon the state of the art machine learning algorithms, which may further extend with the usage of parallel computing approaches like GPU or HADOOP in order to handle massive sizes of satellite images.

ACKNOWLEDGMENT

The authors wish to thank Director, Maharashtra Remote Sensing Applications Centre, Nagpur for indeed support. We also owe our thanks to post graduate research scholar, Ms Neha Mankar for her inputs on few sections of thei research paper. Last but not the least, we would like to express our sincere gratitude to all those who contrinuted directly or indirectly for this research article.

REFERENCES

[1] A. Khobragade, P. Athawale and M. Raguwanshi, Optimization of statistical learning algorithm for crop discrimination using remote sensing data," Advance Computing Conference (IACC), 2015 IEEE International, Banglore, 2015, pp. 570-574.

[2] Anand N. Khobragade, Mukesh M. Raghuwanshi, Contextual soft classification approaches for crops identification using multi-sensory remote sensing data: Machine learning perspective for satellite images, Artificial Intelligence Perspectives and Applications, Volume 347 2015, Proceedings of the 4th Computer Science On-line Conference 2015 (CSOC2015), Vol 1: Artificial Intelligence Perspectives and Applications; DOI: 10.1007/978-3-319-18476-0.

[3] Ying Liu, Bai Zhang, Li-min Wang, Nan Wang, A self-trained semisupervised SVM approach to the remote sensing land cover classification, Computers & Geosciences, Volume 59, September 2013, pp 98-107, ISSN 0098-3004

[4] B. Yekkehkhany, A. Safari, S. Homayouni, M. Hasanlou,ƒA Comparison Study of Different Kernel Functions for SVM-based Classification of Multi-temporal Polarimetry SAR Data,,The 1st ISPRS International Conference on Geospatial Information Research, 15–17 November 2014, Tehran, Iran

[5] Hsiuhan Lexie Yang, Yuhang Zhang, Saurabh Prasad, Melba Crawford, Multiple Kernel Active Learning For Robust Geo-Spatial Image Analysis, IGARSS 2013,publisher IEEE.

[6] Mingmin Chi, RuiFeng, Lorenzo Bruzzone Classification of hyperspectral remote sensing data with primal SVM for small-sized training dataset problemƒ2008 COSPAR. Published by Elsevier Ltd.

[7] Yadong Mu, Gang Hua, Wei Fan, Shih-Fu Chang,ƒHash-SVM: Scalable Kernel Machines for Large-Scale Visual Classification,,Computer Vision and Pattern Recognition (CVPR),

ISSN 2229-5518
2014 IEEE Conference on , pp.979,986, 23-28 June 2014
[8] https://en.wikipedia.org/wiki/Convex_optimization, weblink
[9] Tuia, D.; Matasci, G.; Camps-Valls, G.; Kanevski, M., Learning the relevant image features with multiple kernels,,Geoscience and Remote Sensing Symposium,2009 IEEE International,IGARSS 2009 , vol.2, no., pp.II-65,II-68, 12-17 July 2009
[10] Bor-Chen Kuo, Hsin-Hua Ho, Cheng-Hsuan Li, Chih-Cheng Hung,and Jin-ShiuhTau, A Kernel-Based Feature Selection Method for SVM With RBF Kernel for Hyperspectral Image Classification,,Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of , vol.7, no.1, pp.317,326, Jan. 2014
[11] Kazuhiro Hotta, ƒSupport Vector Machine with Weighted Summation Kernel Obtained by Adaboost,,Video and Signal Based Surveillance, 2006. AVSS '06. IEEE International Conference on , vol., no., pp.11,11, Nov. 2006
[12] Kun Tan, PeijunDu,ƒClassification of hyperspectral image based on morphological profiles and multi-kernel SVM,,Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2010 2nd Workshop on, pp.1,4, 14-16 June 2010.
[13] SabriBoughorbel, Jean-Philippe Tarel, NozhaBoujemaa,ƒConditionally PositiveDefinite Kernels for SVM Based Image RecognitionMultimedia and Expo, 2005.ICME 2005. IEEE International Conference on, pp.113,116, 6-6 July 2005 20
[14] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral images classification," IEEE Trans. Geosci. Remote Sens., vol. 43, no. 6, pp. 1351–1362, Jun. 2005.
[15] Soelaiman, R.; Asfiandy, D.; Purwananto, Y.; Purnomo, M.H.,Weighted kernelfunction implementation for hyperspectral image classification based on Support Vector Machine,,Instrumentation, Communications, Information Technology, andBiomedical Engineering (ICICI-BME), 2009 International Conference on, pp.1,6,23-25 Nov. 2009
[16] Gurram, P.; HeesungKwon,ƒOptimal Sparse Kernel Learning in the Empirical Kernel Feature Space for HyperspectralClassification,,Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of , vol.7, no.4, pp.1217,1226, April 2014
[17] Gurram, P.; HeesungKwon,ƒGeneralized optimal Kernel-Based Ensemble Learning for hyperspectral classification problems ƒGeoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International, pp.4431,4434, 24-29 July 2011
[18] Jiancheng Sun, Xiaohe Li, Yong Yang, Jianguo Luo, and Yaohui Bai. 2012. Scaling the kernel function based on the separating boundary in input space: A data-dependent way for improving the performance of kernel methods. Inf. Sci. 184, 1 (February 2012), 140-154.
DOI=10.1016/j.ins.2011.08.028;http://dx.doi.org/10.1016/j.ins.2011.08.028, ACM
[19] Mingqing Hu; Yiqiang Chen; Kwok, J.T.-Y,Building Sparse Multiple-Kernel SVM Classifiers,Neural Networks, IEEE Transactions on , vol.20, no.5, pp.827,839, May 2009.
[20] Camps-Valls, G.; Gomez-Chova, L.; Munoz-Mari, J.; Rojo-Alvarez, J.L.; MartinezRamon, M,ƒKernel-Based Framework for Multitemporal and Multisource Remote Sensing Data Classification and Change Detection,,Geoscience and Remote Sensing, IEEE Transactions on , vol.46, no.6, pp.1822,1835, June 2008
[21] Hong Cai; YufengWang,ƒGeometrical modification of wavelet SVM kernels and its application in microarray analysis,,Genomic Signal Processing and Statistics (GENSIPS), 2011 IEEE International Workshop on , vol., no., pp.150,153, 4-6 Dec 2011.
[22] Neha Mankar, Anand Khobragade, Mukesh Raghuwanshi, Classification of Remote Sensing Image Using SVM Kernels, IEEE Sponsored World Conference on Futuristic Trends in Research and Innovation for Social Welfare (WCFTR'16),India, 2016, proposed paper.
[23] Zhi He and Junbao Li. 2015. Multiple data-dependent kernel for classification of hyperspectral images. Expert Syst. Appl. 42, 3 (February 2015),1118-1135.DOI=10.1016/j.eswa.2014.09.004 http://dx.doi.org/10.1016/j.eswa.2014.09.004, ACM
[24] Demyanov, L. Backhouse, M. Christie, Geological feature selection in reservoir modelling and history matching with Multiple Kernel Learning, Computers & Geosciences, Available online 26 July 2015, ISSN 0098-3004, http://dx.doi.org/10.1016/j.cageo.2015.07.014, Elsevier publication.
[25] E Li, Daoliang, E Zhao, Chunjiang, A Classification of Remote Sensing Image Based on Improved Compound Kernels of SVM, Book Section, D 2010, @ 978-3-642-12219-4, Computer and Computing Technologies in Agriculture III, V 317, S IFIP Advances in Information and Communication Technology,U http://dx.doi.org/10.1007/978-3-642-12220-0_3, I Springer Berlin Heidelberg,2010
[26] E Ishikawa, Masumi E Doya, Kenji E Miyamoto, Hiroyuki E Yamakawa, Takeshi, Comparison of Local Higher-Order Moment Kernel and Conventional Kernels in SVM for Texture Classification, Book Section D 2008 @ 978-3-540-69154-9 B Neural Information Processing V 4984 S Lecture Notes in Computer Science, U http://dx.doi.org/10.1007/978-3-540-69158-7_88, Springer Berlin Heidelberg ,2008
[27] Compact Fuzzy Multiclass Support Vector Machines & huxia Lu; Pu Shi; Xianhao Liu; Coll. of Math. & Comput. Sci., Hebei Univ., Baoding, Natural Computation, 2008. ICNC '08. Fourth International Conference.
[28] Theodoros Evgenuiu and Massimilliano Pontil, Statistical Learning Theory: a Primer 1998.
[29] Nello Cristianini and John Shawe-Taylor, "An Introduction to Support Vector Machines and Other Kernel-based Learning Methods", Cambridge University Press, 2000.
[30] Tom Mitchell, Machine Learning, McGraw-Hill Computer science series, 1997.
[31] Anil A. Kumar, V.K. Dadhwal, 2010, Entropy-based Fuzzy Classification Parameter Optimization using uncertainty variation across Spatial Resolution, IJRS, pp179-192,2010
IJSER