STUDY OF MIXED KERNEL EFFECT ON CLASSIFICATION ACCURACY USING DENSITY ESTIMATION

Anil Kumar[a*], S. K. Ghosh[b], V. K. Dadhwal[a]

[a]Indian Institute of Remote Sensing, Dehradun, India – (anil, dean)@iirs.gov.in
[b]Indian Institute of Technology Roorkee, India -scnagosh@datainfosys.net

KEYWORDS: Sub-Pixel, Support Vector Machine (SVM), Mixed Kernel Function, Density Estimation, Fuzzy Error Matrix (FERM).

ABSTRACT:

Support Vector Machines (SVMs) are a statistical learning theory based techniques and have been applied to different fields. For the pattern recognition case, SVMs have been used for isolated handwritten digit recognition, object recognition, charmed quark detection, face detection in images and text categorization. SVM have been shown to perform well for density estimation also where the probability distribution function of the feature vector can be inferred from a random sample. In this work SVM has been used for density estimation, and it uses Mean Field (MF) theory for developing an easy and efficient learning procedure for the SVM. In SVM a kernel function determines the characteristic of an SVM. The kernel functions used in SVM are defined as local kernels, global kernels and spectral kernels. In the case of local kernel only the data that are close or in the proximity of each others have an influence on the kernel values. In global kernel samples that are far away from each other still have an influence on the kernel value. A spectral kernel uses the spectral knowledge into SVM classification, which reduces false alarms for thematic classification. In this paper the effect of different mixed kernels generated while taking spectral kernel with local or global kernels have been studied on overall sub-pixel classification accuracy of remote sensing data using Fuzzy Error Matrix (FERM).

1. INTRODUCTION

SVMs are a type of machine learning algorithm that were invented by Vapnik. It has been successfully applied to a wide range of pattern recognition and classification problems including handwriting recognition, and face detection. Support Vector Machines (SVM) is a powerful methodology for solving problems in nonlinear classification, function estimation and density estimation which has also led to many other recent developments in kernel based learning methods in general. SVMs have been introduced within the context of statistical learning theory and structural risk minimization. In SVMs an optimal separating hyperplane between data points of different classes in a (possibly) high dimensional space is calculated. The actual Support Vectors are the points that form the decision boundary between the classes.

Recently SVM have been applied to different fields. For the pattern recognition case, SVMs have been used for isolated handwritten digit recognition (Cortes and Vapnik, 1995; SchÖlkopf, Burges and Vapnik, 1995; SchÖlkopf, Burges and Vapnik, 1996; Burges and SchÖlkopf, 1997), object recognition (Blanz et al., 1996), speaker identification (Schmidt, 1996), charmed quark detection[1], face detection in images (Osuna, Freund and Girosi, 1997), and text categorization (Joachims, 1997). For the regression estimation case, SVMs have been compared on benchmark time series prediction tests (Müller et al., 1997; Mukherjee, Osuna and Girosi, 1997), the Boston housing problem (Drucker et al., 1997), and (on artificial data) on the PET operator inversion problem (Vapnik, Golowich and Smola, 1996). In most of these cases, SVM generalization performance (i.e. error rates on test sets) either matches or is significantly better than that of competing methods. Regarding extensions, the basic SVMs contain no prior knowledge of the problem (for example, a large class of SVMs for the image recognition problem would give the same results if the pixels were first permuted randomly (with each image suffering the same permutation), an act of vandalism that would leave the best performing neural networks severely handicapped) and much work has been done on incorporating prior knowledge into SVMs (SchÖlkopf, Burges and Vapnik, 1996; SchÖlkopf et al., 1998a; Burges, 1998). Although SVMs have good generalization performance, they can be abysmally slow in test phase, a problem addressed in (Burges, 1996; Osuna and Girosi, 1998). Recent work has generalized the basic ideas (Smola, SchÖlkopf and Müller, 1998a; Smola and SchÖlkopf, 1998), shown connections to regularization theory (Smola, SchÖlkopf and Müller, 1998b; Girosi, 1998; Wahba, 1998), and shown how SVM ideas can be incorporated in a wide range of other algorithms (SchÖlkopf, Smola and Müller, 1998b; SchÖlkopf et al, 1998c) (Christopher et al., 1998).

In this work SVMs had been used for density estimation. In this case Mean Field (MF) theory had been used for developing an easy and efficient learning procedure for the SVM. But the traditional formulation of the SVM density estimation decomposes the parameters of the problem into a quadratic optimization, which is solved using standard optimization techniques. In this paper the effect of different kernels while generating density estimation using SVM have been studied with respect to overall sub-pixel classification accuracy of multi-spectral data. This work was done using SMIC (Sub-Pixel Multi-Spectral Image Classifier) System (Kumar et al., 2005).

2. KERNELS IN SVM

SVMs are designed to solve two-class problems. Two approaches can be used for a M-class problem. One approach is called one against all; in this M classifiers are iteratively applied on each class against all the others. Other is called one

* Corresponding author.

against one; $\frac{M(M-1)}{2}$ classifiers are applied on each pair of classes, the most often computed label is kept for each vector. The kernel function is constructed by SVM algorithm to map the training data into a higher dimensional space when the linear separation is impossible in the original one. SVM can be generalized to compute nonlinear decision surfaces. The method consists in projecting the data in a higher space where they are considered to become linearly separable. SVM applied in this space lead to the determination of nonlinear surfaces in the original space. Actually, the projection can be simulated using a kernel method (Grégoire et al, 2003).

Every function $K(\cdot,\cdot)$ that satisfies mercer's conditions may be considered as an eligible kernel. The Mercer's conditions state as:

$$\forall g(\cdot)\varepsilon\, L^2(\Re^n) \text{ so that } \int g(x)^2\, dx$$

is finite.

then $\int K(x,y)g(x)g(y)dxdy \geq 0.$

A great number of kernels exist and it is difficult to explain their individual characteristics. The kernels used in work are known as local kernels, global kernels and spectral kernels, which are mentioned as follows;

Local kernels: Only the data that are close or in the proximity of each other's have an influence on the kernel values. Basically, all kernels that are based on a distance function are local kernels. Examples of typical local kernels are;

Gaussian:

$$\mathrm{K}(x,x_i) = \exp(-0.5(x-x_i)A^{-1}(x-x_i)^T$$

(Refaat et al, 2004)

Were A have three following norms;

$\mathrm{A} = \mathrm{I}$	Euclidean Norm
$A = D_j^{-1}$	Diagonal Norm
$A = C_j^{-1}$	Mahalonobis Norm

Radial basis: $\mathrm{K}(x, x_i) = \exp(-\| x - x_i \|^2)$

KMOD: $\mathrm{K}(x,x_i) = \exp\left(\frac{1}{1+\| x-x_i \|^2}\right)$ -1

Inverse Multiquadric: $\mathrm{K}(x,x_i) = \frac{1}{\sqrt{(\| x-x_i \|^2 + 1)}}$

Global kernels: Samples that are far away from each others still have an influence on the kernel value. All kernels based on the dot product are global. Examples of typical global kernels are;

Linear: $\mathrm{K}(x,x_i) = x.x_i$

Polynomial: $\mathrm{K}(x,x_i) = (x.x_i + 1)^p$

Sigmoid: $\mathrm{K}(x,x_i) = \tanh(x.x_i + 1)$

Spectral kernels: The local kernels are based on a quadratic distance evaluation between two samples. In order to fit hyperspectral point of view, it is of interest to consider new criteria that take into consideration spectral signature concept. Spectral angle (SA) $\alpha(x,x_i)$ is defined in order to measure the spectral difference between x and x_i while being robust to differences of the overall energy (e.g. illumination, shadows etc.) (Grégoire et al, 2003).

Spectral angle (SA): $\alpha(x,x_i) = \arccos\left(\frac{x.x_i}{\| x \| \; \| x_i \|}\right)$

As in remote sensing data multi-spectral images are sharpen while fussing multi-spectral image with panchromatic image. Same in the case of kernel function, mixture of kernels can be used (Grégoire et al, 2003). Linear mixture of kernels can fit the dual characteristics; characteristics of dot product or Euclidian distance and also characteristic of spectral angle. Mixture of kernels may be defined as:

$$\mathrm{K}(x,x_i) = \mu\, \mathrm{K}_a(x,x_i) + (1-\mu)\mathrm{K}_b(x,x_i)$$

where $\mathrm{K}_a(x,x_i)$ and $\mathrm{K}_b(x,x_i)$ are two kernels (e.g. local, global and spectral angle). Since $\mathrm{K}_a(x,x_i)$ and $\mathrm{K}_b(x,x_i)$ satisfy Mercer's conditions, all linear combinations are eligible for kernels. In this work $\mathrm{K}_a(x,x_i)$ kernel has been taken any local or global kernel and $\mathrm{K}_b(x,x_i)$ kernel has been taken as Spectral kernel.

3. MULTI-SPECTRAL IMAGE

A UTM rectified LISS-III image from Resourcesat –1, (IRS-P6) satellite acquired in four bands have been used. The image was acquired in 2003 and covers the rural area of Dehradun District, of Uttaranchal State, India. Approximately, 30% of the area in the image selected is covered by reserved forest, 30% by agriculture land, 15% by barren land, 15% sand, and 10% by river water. The size of the image is 250 × 296 pixels, with spatial resolution of 23.5 m. The five classes of interest, namely forest, water, agriculture, barren, and sand have been used for this study work (Figure 1).

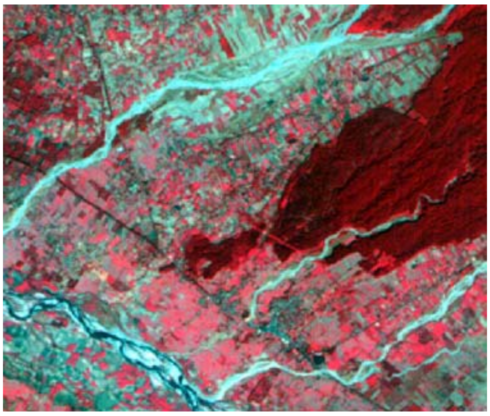

Figure 1. False Colour Composite from the Resource Sat-1 LISS-III multi-spectral image.

4. KERNEL FUNCTION VERSES OVERALL ACCURACY

The effect of different kernel functions on sub-pixel classification of LISS-III image from Resourcesat –1, (IRS-P6) satellite were studied while using density estimation algorithm based on Support Vector Machine approach for sub-pixel classification. The learning parameters for Support Vector Machine approach were kept constant for all the kernel functions. The training as well as testing data used for supervised approach was >10n, were n is dimension of data used. Separate data were used at training as well as at testing stage. At testing stage 500 samples were taken for overall accuracy assessment of sub-pixel output. The effect of different kernel functions were observed on sub-pixel classification output using Fuzzy Error Matrix (Binaghi et al., 1999). The overall accuracy of sub-pixel classification, obtained while using different combinations of kernel functions in Support Vector Machine approach are mentioned in table 1;

Sl. No.	Mixed Kernel Function		Overall Accuracy (%)			
	K_a	K_b	$\mu = 0.1$	$\mu = 0.2$	$\mu = 0.3$	$\mu = 0.5$
1	Gaussian with Euclidean Norm	Spectral Kernel	94.56	91.22	93.41	91.58
2	Gaussian with Mahalonobis Norm	Spectral Kernel	90.85	93.70	91.25	93.79
3	Gaussian with Diagonal Norm	Spectral Kernel	88.23	93.36	93.19	90.85
4	Radial basis (σ = 0.25)	Spectral Kernel	92.05	93.96	94.57	91.98
5	KMOD	Spectral Kernel	91.23	91.8	93.75	92.86
6	Inverse Multiquadric	Spectral Kernel	92.26	93.52	94.12	92.82
7	Linear	Spectral Kernel	91.89	93.30	93.61	93.06
8	Polynomial (1st order)	Spectral Kernel	92.56	91.06	94.44	91.54
9	Sigmoid	Spectral Kernel	90.74	90.19	91.95	94.27

Table 1. Overall Accuracy while using different Mixed Kernel functions.

5. CONCLUSION

Basically, all kernels that are based on a distance function are local kernels and in local kernels the data that are close or in the proximity of each other's have an influence on the kernel values. But in global kernels samples that are far away from each other still have an influence on the kernel value. All kernels based on the dot product are global. In spectral kernel, Spectral angle (SA) $\alpha(x, x_i)$ is defined in order to measure the spectral difference between x and x_i while being robust to differences of the overall energy (e.g. illumination, shadows etc.). To fit the dual point of view: similarity according to the dot product or Euclidian distance and also, similarity according to the spectral shape (SA), Mixture of kernels have been used in this study.

Table: 1, shown in section 4, shows that overall sub-pixel classification accuracy of multi-spectral remote sassing data varies while using different mixture of kernel functions in SVM. The maximum mixtures of kernels are giving maximum sub-pixel classification accuracy when μ in mixture of kernel used is of range 0.3. Some mixture of kernels gives good accuracy when μ =0.5 value used in mixture of kernel.

REFERENCES

1. Kumar A., Ghosh S. K., Dadhwal V. K., 2005, Advanced Supervised Sub-Pixel Multi-Spectral Image Classifier, Map Asia 2005 conference, Jakarta. Indonesia.

2. Binaghi, E., Brivio, P. A., Chessi, P., and Rampini, A., 1999, A fuzzy Set based Accuracy Assessment of Soft Classification, Pattern Recognition letters, 20, 935-948.

3. Blanz, V., SchÖlkopf, B., Bülthoff, H., Burges, C., Vapnik, V. and Vetter, T. 1996, Comparison of view–based object recognition algorithms using realistic 3d models. In C. von der Malsburg, W. von Seelen, J. C. Vorbrüggen, and B. Sendhoff, editors, *Artificial Neural Networks—ICANN'96*, pages 251 – 256, Berlin, Springer Lecture Notes in Computer Science, Vol. 1112.

4. Burges, C.J.C. 1996, Simplified support vector decision rules. In Lorenza Saitta, editor, *Proceedings of the Thirteenth International Conference on Machine Learning*, pages 71–77, Bari, Italy, Morgan Kaufman.

5. Burges, C.J.C. and SchÖlkopf, B. 1997, Improving the accuracy and speed of support vector learning machines. In M. Mozer, M. Jordan, and T. Petsche, editors, *Advances in Neural Information Processing Systems 9*, pages 375–381, Cambridge, MA, MIT Press.

6. Burges, C. J. C. 1998, Geometry and invariance in kernel based methods. In *Advances in Kernel Methods - Support Vector Learning,* Bernhard SchÖlkopf, Christopher J.C. Burges and Alexander J. Smola (eds.), MIT Press, Cambridge, MA.

7. Cortes, C. and Vapnik, V. 1995, Support vector networks. *Machine Learning*, 20:273–297.

8. Drucker, H., Burges, C.J.C., Kaufman, L., Smola, A. and Vapnik, V. 1997, Support vector regression machines. *Advances in Neural Information Processing Systems*, 9:155–161.

9. Grégoire Mercier and Marc Lennon, 2003, Support Vector Machines for Hyperspectral Image Classification with Spectral-based kernels, IGARSS.

10. Girosi, F. 1998, An equivalence between sparse approximation and support vector machines. *Neural Computation; CBCL AI Memo 1606, MIT.*

11. Joachims, T. 1997, Text categorization with support vector machines. Technical report, LS VIII Number 23, University of Dortmund. ftp://ftp-ai.informatik.uni-dortmund.de/pub/Reports/report23.ps.Z.

12. Mukherjee, S., Osuna, E. and Girosi, F. 1997, Nonlinear prediction of chaotic time series using a support vector machine. In *Proceedings of the IEEE Workshop on Neural Networks for Signal Processing 7*, pages 511–519, Amelia Island, FL.

13. Müller, K. R., Smola, A., Rätsch, G., SchÖlkopf, B., Kohlmorgen, J. and Vapnik, V. 1997, Predicting time series with support vector machines. In *Proceedings, International Conference on Artificial Neural Networks*, page 999. Springer Lecture Notes in Computer Science.

14. Osuna, E., Freund, R. and Girosi, F. 1997, An improved training algorithm for support vector machines. In *Proceedings of the 1997 IEEE Workshop on Neural Networks for Signal Processing, Eds. J. Principe, L. Giles, N. Morgan, E. Wilson*, pages 276 – 285, Amelia Island, FL.

15. Osuna, E. and Girosi. F. 1998, Reducing the run-time complexity of support vector machines. In *International Conference on Pattern Recognition (submitted).*

16. Refaat M Mohamed, and Aly A Farag, 2004, Mean Field Theory for Density Estimation Using Support Vector Machines, Computer Vision and Image Processing Laboratory, University of Louisville, Louisville, KY, 40292.

17. SchÖlkopf, B., Burges, C. and Vapnik, V. 1995, Extracting support data for a given task. In U. M. Fayyad and R. Uthurusamy, editors, *Proceedings, First International Conference on Knowledge Discovery & Data Mining*. AAAI Press, Menlo Park, CA.

18. SchÖlkopf, B., Burges, C. and Vapnik, V. 1996, Incorporating invariances in support vector learning machines. In C. von der Malsburg, W. von Seelen, J. C. Vorbrüggen, and B. Sendhoff, editors, *Artificial Neural Networks — ICANN'96*, pages 47 – 52, Berlin, Springer Lecture Notes in Computer Science, Vol. 1112.

19. SchÖlkopf, B., Smola, A., Müller, K. R., Burges, C.J.C. and Vapnik, V. 1998, Support vector methods in learning and feature extraction. In *Ninth Australian Congress on Neural Networks*.

20. SchÖlkopf, B., Smola, A. and Müller, K. R. 1998, Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*. In press.

21. Schmidt, M. 1996, Identifying speaker with support vector networks. In *Interface 96 Proceedings*, Sydney.

22. Smola, A. J., and SchÖlkopf, B. 1998, On a kernel-based method for pattern recognition, regression, approximation and operator inversion. *Algorithmica*.

23. Smola, A. J., SchÖlkopf, B. and Müller, K. R. 1998, General cost functions for support vector regression. In *Ninth Australian Congress on Neural Networks.*

24. Smola, A. J., SchÖlkopf, B. and Müller, K. R. 1998, The connection between regularization operators and support vector kernels. *Neural Networks.*

25. Vapnik, V., Golowich, S. and Smola, 1996, A. Support vector method for function approximation, regression estimation, and signal processing. *Advances in Neural Information Processing Systems*, 9:281–287.

26. Wahba G. 1998, Support vector machines, reproducing kernel hilbert spaces and the randomized gacv. In *Advances in Kernel Methods - Support Vector Learning*, Bernhard SchÖlkopf, Christopher J.C. Burges and Alexander J. Smola (eds.), MIT Press, Cambridge, MA.