© EARSeL and Warsaw University, Warsaw 2005. *Proceedings of 4th EARSeL Workshop on Imaging Spectroscopy. New quality in environmental studies*.
Zagajewski B., Sobczak M., Wrzesień M., (eds)

MULTI-CLASS SUPPORT VECTOR MACHINE CLASSIFICATION FOR HYPERSPECTRAL DATA

Xiuping Jia

University of New South Wales, University College, School of Information Technology and Electrical Engineering, Canberra, Australia; x-jia@adfa.edu.au

ABSTRACT

A progressive two-class decision classifier (pTCDC) was developed for hyperspectral data mapping to achieve maximum class separations between each class pair. In this paper, pTCDC is tested further by comparing it with other possible ways of converting multiclass to two-class classification including one-against-all and one-to-one methods used in implementing the newly developed support vector machine classifier for remote sensing data. Experiments carried out using an AVIRIS data set are presented and the results demonstrate that pTCDC is more efficient than that of one-to-one structure and more reliable than one-again-all method.

INTRODUCTION

Parametric supervised classification techniques, such as Gaussian maximum likelihood, have been widely adopted for multispectral remotes sensing image classification (1). However, when the number of spectral bands is high, class data modeling presents great challenges in order to reflect the class data's real distribution in a high dimensional feature space with a limited number of training samples. Band selection and feature extraction can be introduced as preprocessing to solve the problem. However, the selected subset of data may not be optimal for each class pair in terms of their class separability. Multistage classification has been suggested which provides flexibility in features to use and a decision rule to use at each stage. Binary decision tree is preferred structure among all the mutlistage classification due to its regularity. On the other hand, non-parametric classification method is an alternative to avoid the difficulties in estimate of class data's statistical parameters. One of the non-parametric pixel labeling algorithm is k nearest neighbors (k-NN), which bypasses density function estimation and goes directly to a decision rule (2).

In recent years, support vector machines (SVM) (3) have been introduced to remote sensing image classification (4). It is a non-parametric approach aiming at finding a decision hyperplane by maximizing the margin between the separating plane and the data. The training is performed on two classes at a time. A direct multiclass SVM training (all-together) was proposed, however, it was found low performance than converting multiclass problem to two-class classification (5). Two methods for applying Bi-SVM to M-class classification are commonly adopted. They are One-against-All, which constructs SVM decision boundary to separate each class from the rest in turn, and One–to–One, which constructs SVM for each class pair.

In this paper, a progressive two-class decision classifier (pTCDC, (6)) (or a Directed Acyclic Graph (DAG) (7)) developed earlier is tested and compared to One-against-All and One-to-One methods. The next section SVM classification scheme and pTCDC will be presented, followed by experiments and discussions.

METHODS

Figure 1 illustrates the idea of SVM classifier. It shows a scatter plot of a two-dimensional training data for two classes. SVMs aim at finding an optimal decision hyperplane that makes the distance to the closest vectors - support vectors, in each side maximum.

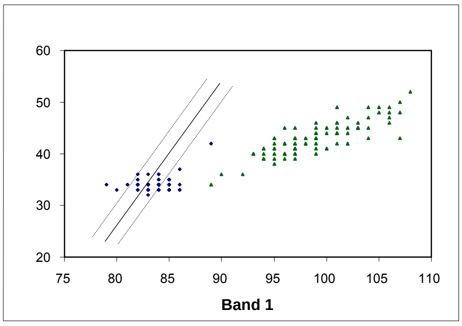

Figure 1: An illustration of SVM classifier. The training samples on the dotted lines are called support vectors; the solid line is the optimum separating hyperplane.

The separating plane is obtained by using constrained optimization techniques and details can be found in (8). When the class data are not linearly separable, a data transformation using a kernel function can be applied to map the data to a higher dimensional space. A kernel function can be (9)

Linear: $K(\mathbf{x}_i,\mathbf{x}_j)=\mathbf{x}_i^t\mathbf{x}_j$

Polynomial: $K(\mathbf{x}_i,\mathbf{x}_j)=(\gamma\mathbf{x}_i^t\mathbf{x}_j+r)^d,\ \gamma>0$

Radial basis function (RBF): $K(\mathbf{x}_i,\mathbf{x}_j)=\exp(-\gamma\left\|\mathbf{x}_i-\mathbf{x}_j\right\|^2)\quad \gamma>0$

Sigmoid: $K(\mathbf{x}_i,\mathbf{x}_j)=\tanh(\gamma\mathbf{x}_i^t\mathbf{x}_j+r)$

where $\mathbf{x}$ is a row of column vectors, each vector is the data of a training sample. γ, *r*, and *d* are kernel parameters. When RBF is used, γ can be determined together with C (the penalty parameter of the error term in the optimization) via cross validation. If the parameters are selected properly, it can improve classification accuracy (10, 11 and 12).

The training is performed for two classes at a time. The multiclass problem needs to be handled with appropriate multiclass to two-class conversion. Two methods for applying SVM to M-class classification are commonly adopted:

One-against-all: Construct SVM to classify every class against the rest in turn, resulting in M machines. For a testing pixel, the majority vote from the number of machines determines the label for the pixel.

One–to–one: Construct SVM for each pair of classes, resulting in M(M-1)/2 machines. For a testing pixel, the majority vote from the number of machines determines the label for the pixel (This is also called a Max Wins algorithm).

Majority voting strategy is also called a Max Wins algorithm (5). Instead of using majority voting, coding matrix together with Hamming distance may be used.

The Progressive Two-Class Decision Classifier (7) or a Directed Acyclic Graph (DAG) (8) developed earlier for hyperspectral image classification is an alternative multicasts to two-class conversion method. Figure 2 shows the structure of the scheme. $\mathbf{X}$ is an N-dimensional pixel vector. Suppose there are six categories, represented by $\omega_a,\omega_b,\omega_c,\omega_d,\omega_e$ and ω_f. The scheme focuses on one class pair at a time (at a node). The function of the very first layer is to check the potential membership of $\mathbf{X}$ to class ω_a and class ω_b and the vector is classified temporarily as either class ω_a or class ω_b using the decision rule, D_{ab}, which can be the results from SVM training. Class ω_b will be rejected for further consideration for those vectors labeled into the ω_a category, and class ω_a is rejected for further consideration for those vectors labeled into the ω_b category. At the second layer, there are two nodes, and two new class pairs - ω_a and ω_c for the left side node and ω_b and ω_c for the right side node - are considered, respectively. This process continues until a pure class labeling has been reached at the last layer, which is the final assignment. This process can be interpreted in a different way by following Fig. 2 in the vertical direction, instead of the horizontal. Each column corresponds to a particular class and a pixel is examined progressively as belonging to that class or not. The structure can be easily extended for cases of more than six classes.

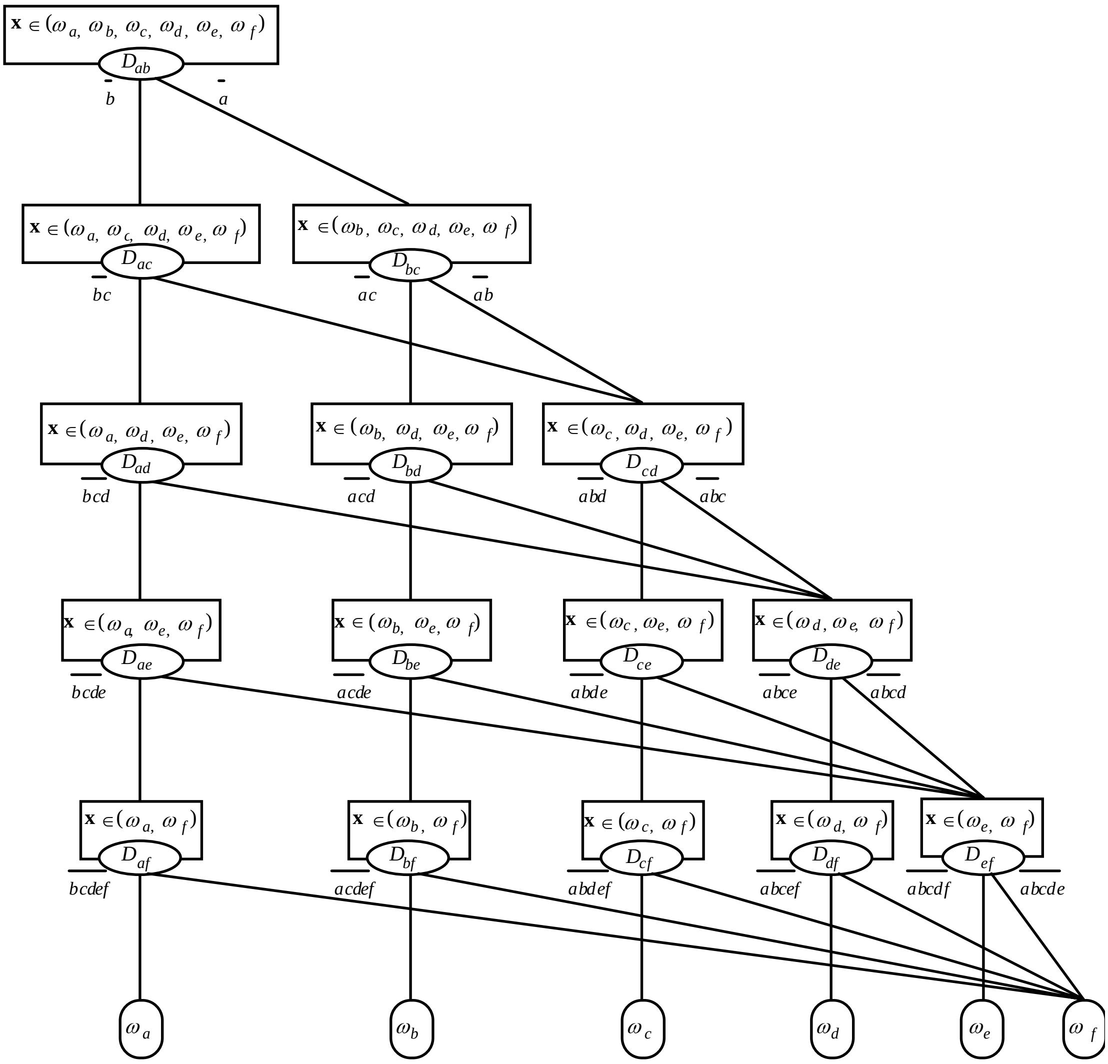

Fig. 2. A schematic chart for progressive two-class decision classifier with 6 classes: $\omega_a, \omega_b, \omega_c, \omega_d, \omega_e$ **and** ω_f. $\mathbf{x}$*: a pixel vector.* ω_i*: class* i*;* D_{ij}*: decision rule used for separating class* i *and* j*.*

To use pTCDC together with SVM, M(M-1)/2 SVMs need to be trained for all the class pairs. For a testing pixel, M-1 machines will be used to label it. The scheme performance in terms of classification accuracy is identical to one-to-one methods and higher than one-against-all method since one-against-all may not provide the best separation. The computational load of pTCDC is greatly reduced comparing with one-to-one structure due to the fewer decisions to combine and without the need for majority voting. These are tested in the following section.

RESULTS

AVIRIS data covering an area of mixed agriculture and forestry in Northwestern Indiana, USA, was used in the following. The data was recorded in June, 1992 with 220 bands. Water absorption bands, 104 to 108 and 150 to 162, were removed, leaving 202 bands for analysis. 4 classes, -'Soybean', 'Corn', 'Grass' and 'Hay'- were selected. Details are given in Fig. 3 and Table 1. A traditional principal components transformation was applied to the data; the first 40 new features, which contained 99.93% of total variance of the original data, were used for this experiment.

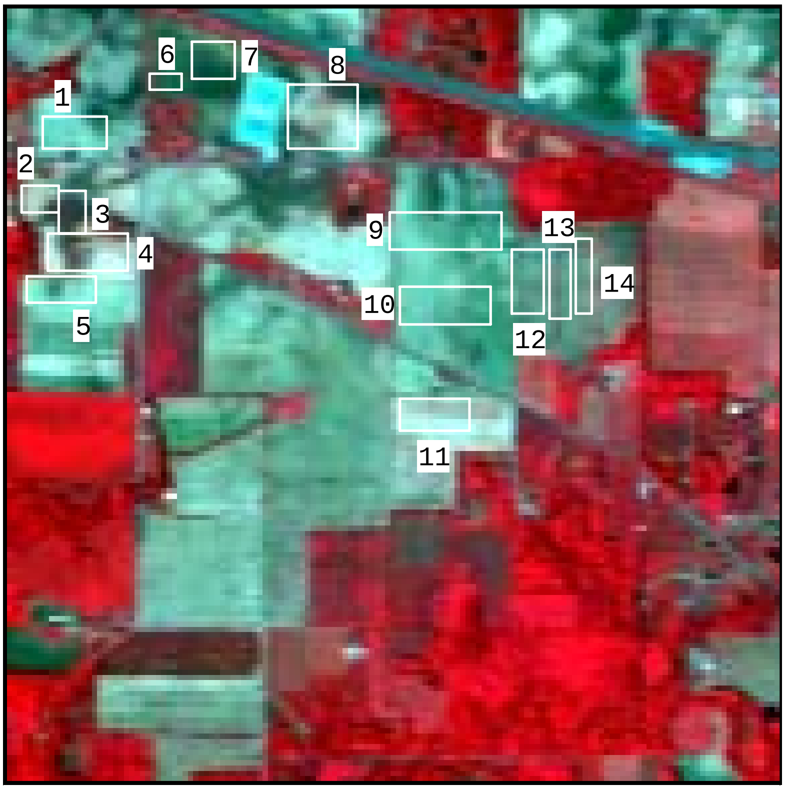

Figure 3: The AVIRIS image with training and testing fields selections.

Table 1: Training and testing data

Class Number	Class Name	Number of Training Pixels	Number of Testing Pixels
1	Corn no-till	130	144
2	Corn clear-till	105	75
3	Soybean no-till	137	189
4	Soybean clear-till	121	156

The following parameters were used for the SVM training with the OSU Support Vector Machines Toolbox (13)

Kernel function: linear

Cost of constrains violation: 1

Tolerance of termination criterion: 0.001

All the 6 class pairs were considered separately for one-to-one and pTCDC. Also, each class against rest of three classes was trained to implement one-to-all scheme. The classification accuracies obtained are given in Table 2. It can be seen that one-to-one training is more reliable. However, one-to-one requires higher computational load. The pTCDC has the same performance as one-to-one in term of classification accuracy with lower computational load. The details are given in Table 3.

Table 2: Classification Accuracy

	SVM	
	One-to-One	One-to-All
No. of Features	40	40
Training Data	100%	98.0%
Testing Data	60.3%	49.1%

Table 3: Computational load for training and labeling comparison

	One-to-One		One-to-All		pDCTC	
	4 Classes	M Classes	4 Classes	M Classes	4 Classes	M Classes
No. of SVM	6	M(M-1)/2	4	M[1]	6	M (M-1)/2
No. of Labeling (Per Pixel)	6	M(M-1)/2	4	M[1]	3	M-1
Majority Vote	Yes		Yes		No	

[1] The training time for each One-to-All machine is longer than that for One-to-One, since the larger number of training samples involved.

CONCLUSIONS

Support vector machine classification is a method to find the best separation hyperplane between two categories. Its performance can be enhanced with the Progressive Two-Class Decision Classification scheme (pTCDC).

pTCDC is more efficient than that of one-to-one structure since it reduces the number of machines by the ratio of M/2 (from M(M-1)/2 to M-1, M is the number of classes), and remove the need of majority voting for each pixel labelling.

pTCDC is more reliable than one-again-all method, since a good separation between one class and the rest of mixed class is difficult to achieve comparing between two classes.

SVMs are sensitive to the support vectors, therefore, training data are expected to have good quality. Moreover, the parameters for the kernel function to use require extensive training before a set of classifiers can be set up.

ACKNOWLEDGEMENTS

The author thanks J. Ma, Y. Zhao and S. Ahalt of The Ohio State University for their OSU SVM Classifier Matlab Toolbox (version 3.00). The author also thanks Dr. D. Landgrebe of the Purdue University for providing the AVIRIS data set and the MultiSpec software package.

References

(1) J. A. Richards and X. Jia (1999). *Remote Sensing Digital Image Analysis*, 3rd Ed., Springer Verlag.

(2) B. V. Dasarathy (Editor) (1990). *Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques,* IEEE Computer Society Press, Los Alamitos, California.

(3) V.N. Vapnik (1998). *Statistical Learning Theory,* New York: Wiley.

(4) J.A. Gualtieri and R.F. Cromp (2000). Support vector machines for classification of hyperspectral remote-sensing images, *Proc. IGARSS'00*, Honolulu, 2000, pp.813-815.

(5) C-W. Hsu and C-J. Lin (2002). A comparison of methods for multiclass support vector machines, *IEEE Trans. On Neural Networks*, Vol. 13, No. 2, pp. 415-425.

(6) X. Jia and J.A. Richards (1998). Progressive two-class decision classifier for optimization class discriminations, *Remote Sensing of Environment,* Vol. 63, No. 3, pp. 289-297.

(7) J. C. Platt, N. Cristianini and J. Shawe-Taylor (2000). Large margin DAGs for multiclass classification, *Advances in Neural Information Processing Systems,* vol. 12, S.A. Solla, T.K. Leen and K.-R. Muller (eds.), 547-553, MIT Press.

(8) C.J. C. Burges (1998). A tutorial on support vector machines for pattern recognition, *Knowledge Discovery and Data Mining,* Vol. 2, 1998.

(9) C-W. Hsu, C-C. Chang and C-J, Lin. A practical guide to support vector classification, http://www.csie.ntu.edu.tw/~cjlin/libsvm/

(10) J. Zhang, Y. Zhang and T. Zhou (2001). Classification of hyperspectral data using support vector machine, *Proceedings, International Conference on Image Processing,* Volume 1, 7-10 Oct. , pp. 882 - 885.

(11) F. Melgani and L. Bruzzone (2004). Classification of hyperspectral remote sensing images with support vector machines*, IEEE Transactions on Geoscience and Remote Sensing*, Vol. 42, No. 8, pp.1778 -1790 .

(12) G.M. Foody and A. Mathur (2004). A relative evaluation of multiclass image classification by support vector machines, *, IEEE Transactions on Geoscience and Remote Sensing*, Vol. 42, No. 6, pp. 1335 -1343.

(13) J. Ma, Y. Zhao and S. Ahalt, OSU SVM Classifier Matlab Toolbox (version 3.00) http://www.ece.osu.edu/~maj/osu_svm/