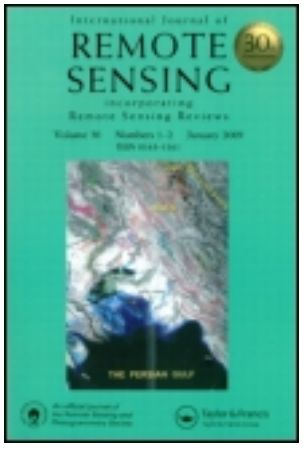

International Journal of Remote Sensing

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tres20

Random forest classifier for remote sensing classification

M. Pal [a]

[a] National Institute of Technology, Department of Civil Engineering, Kurukshetra, 136119, Haryana, India E-mail:

Version of record first published: 22 Feb 2007.

To cite this article: M. Pal (2005): Random forest classifier for remote sensing classification, International Journal of Remote Sensing, 26:1, 217-222

To link to this article: http://dx.doi.org/10.1080/01431160412331269698

International Journal of Remote Sensing
Vol. 26, No. 1, 10 January 2005, 217–222

Taylor & Francis
Taylor & Francis Group

Random forest classifier for remote sensing classification

M. PAL
Department of Civil Engineering, National Institute of Technology, Kurukshetra, 136119, Haryana, India; Email: mpce_pal@yahoo.co.uk

(*Received 1 October 2003; in final form 1 May 2004*)

Growing an ensemble of decision trees and allowing them to vote for the most popular class produced a significant increase in classification accuracy for land cover classification. The objective of this study is to present results obtained with the random forest classifier and to compare its performance with the support vector machines (SVMs) in terms of classification accuracy, training time and user defined parameters. Landsat Enhanced Thematic Mapper Plus (ETM+) data of an area in the UK with seven different land covers were used. Results from this study suggest that the random forest classifier performs equally well to SVMs in terms of classification accuracy and training time. This study also concludes that the number of user-defined parameters required by random forest classifiers is less than the number required for SVMs and easier to define.

1. Introduction

In recent years, a number of papers reported the use of a combination of multiple classifiers to produce a single classification in the remote sensing literature (Giacinto and Roli 1997, Breim *et al.* 2002). The resulting classifier, referred to as an ensemble classifier, is generally found to be more accurate than any of the individual classifiers making up the ensemble (Dietterich 2002). An ensemble classifier combines the decision of a set of classifiers by weighted or unweighted voting to classify unknown examples. Studies using boosting (Freund and Schapire 1996) with a decision tree as a base classifier have reported a significant increase in classification accuracy for land cover classification studies (Friedl *et al.* 1999, Muchoney *et al.* 2000, Pal and Mather 2003a). The aim of this Letter is to discuss the results obtained using the random forest classifier as proposed by Breiman (1999). This classifier involves choosing a set of features randomly and creating a classifier with a bootstrapped sample of the training data. A large number of trees (classifiers) are generated in this way and finally unweighted voting is used to assign an unknown pixel to a class. Further, the performance of the random forest classifier is compared with support vector machines in terms of classification accuracy, training time and user-defined parameters.

2. Classification algorithms

The random forest classifier, which is a tree-based classifier, and support vector machines, based on maximizing the margin between two different classes, are used in this study.

International Journal of Remote Sensing
ISSN 0143-1161 print/ISSN 1366-5901 online © 2005 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/01431160412331269698

2.1 *Random forest classifier*

The random forest classifier consists of a combination of tree classifiers where each classifier is generated using a random vector sampled independently from the input vector, and each tree casts a unit vote for the most popular class to classify an input vector (Breiman 1999). The random forest classifier used for this study consists of using randomly selected features or a combination of features at each node to grow a tree. Bagging, a method to generate a training dataset by randomly drawing with replacement N examples, where N is the size of the original training set (Breiman 1996), was used for each feature/feature combination selected. Any examples (pixels) are classified by taking the most popular voted class from all the tree predictors in the forest (Breiman 1999). Design of a decision tree required the choice of an attribute selection measure and a pruning method. There are many approaches to the selection of attributes used for decision tree induction and most approaches assign a quality measure directly to the attribute. The most frequently used attribute selection measures in decision tree induction are the Information Gain Ratio criterion (Quinlan 1993) and the Gini Index (Breiman *et al.* 1984). The random forest classifier uses the Gini Index as an attribute selection measure, which measures the impurity of an attribute with respect to the classes. For a given training set T, selecting one case (pixel) at random and saying that it belongs to some class C_i, the Gini index can be written as:

$$\sum \sum_{j \neq i} (f(C_i, T)/|T|)\big(f\big(C_j, T\big)/|T|\big) \tag{1}$$

where $f(C_i, T)/|T|$ is the probability that the selected case belongs to class C_i.

Each time a tree is grown to the maximum depth on new training data using a combination of features. These fully grown trees are not pruned. This is one of the major advantages of the random forest classifier over other decision tree methods like the one proposed by Quinlan (1993). The studies suggest that the choice of the pruning methods, and not the attribute selection measures, affect the performance of tree based classifiers (Mingers 1989, Pal and Mather 2003a). Breiman (1999) suggests that as the number of trees increases, the generalization error always converges even without pruning the tree and overfitting is not a problem because of the Strong Law of Large Numbers (Feller 1968). The number of features used at each node to generate a tree and the number of trees to be grown are two user-defined parameters required to generate a random forest classifier. At each node, only selected features are searched for the best split. Thus, the random forest classifier consists of N trees, where N is the number of trees to be grown, which can be any value defined by the user. To classify a new dataset, each case of the datasets is passed down to each of the N trees. The forest chooses a class having the most out of N votes, for that case.

2.2 *Support Vector Machines (SVMs)*

SVMs are based on statistical learning theory and have the aim of determining the location of decision boundaries that produce the optimal separation of classes (Vapnik 1995). In a two-class pattern recognition problem where classes are linearly separable, the SVMs select the one linear decision boundary that leaves the greatest margin between the two classes. The margin is defined as the sum of the distances to the hyperplane from the closest points of the two classes (Vapnik 1995). This

problem of maximizing the margin can be solved using standard Quadratic Programming (QP) optimization techniques. The data points that are closest to the hyperplane are used to measure the margin. Therefore, these data points are termed 'support vectors' and are always small in number (Vapnik 1995).

If the two classes are not linearly separable, the SVMs try to find the hyperplane that maximizes the margin, while at the same time, minimizing a quantity proportional to the number of misclassification errors. The trade-off between margin and misclassification error is controlled by a positive user-defined parameter C (Cortes and Vapnik 1995). SVMs can also be extended to handle nonlinear decision surfaces. Boser *et al.* (1992) proposed a method for projecting the input data into a high-dimensional feature space through some nonlinear mapping, and formulating a linear classification problem in that feature space. Kernel functions are used to reduce the computational cost of dealing with high-dimensional feature space (Vapnik 1995).

SVMs were initially designed for binary (two-class) problems. When dealing with multiple classes, an appropriate multi-class method is needed. Techniques such as 'one against one' and the 'one against the rest' are in frequent use for the multi-class problems (Vapnik 1995, Cristianini and Shawe-Taylor 2000).

3. Application

For this study, Landsat-7 Enhanced Thematic Mapper (ETM+) data (19 June 2000) of an agricultural area near Littleport, Cambridgeshire, UK was used. An area of 307-pixel (columns) by 330-pixel (rows) covering the area of interest was used for this study. The classification problem involved the identification of seven land cover types (wheat, potato, sugar beet, onion, peas, lettuce and beans). Field data printouts for the relevant crop season were collected from farmers and their representative agencies. The other areas were surveyed on the ground to prepare the ground reference image. A total of 4737 pixels were selected for all seven classes by using equalized random sampling. Pixels were then divided into two parts so as to remove any possible bias caused by using the same pixels for training and testing the classifiers. A total of 2700 training and 2037 test pixels were used (table 1).

A number of trials were carried out to select the user-defined parameters for random forest classifier (figures 1 and 2). For this study three numbers of features at each node and a total of 100 trees were used.

A set of user-defined parameters is required to design SVMs. Choice of kernel, kernel specific parameters and the value of the regularization parameter (C) are found to influence the classification accuracy achieved by SVMs. A radial basis

Table 1. Number of pixels used for different classes during training and testing.

Class	Number assigned to different classes	Training pixels	Test pixels
Wheat	1	400	300
Sugar beat	2	400	300
Potato	3	400	300
Onion	4	400	300
Peas	5	400	300
Lettuce	6	400	300
Beans	7	300	237

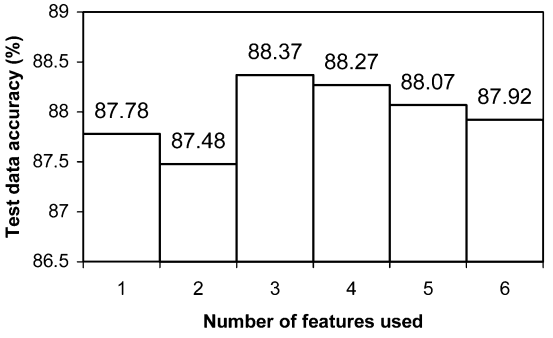

Figure 1. Variation in classification accuracy with a varying number of features using 100 trees.

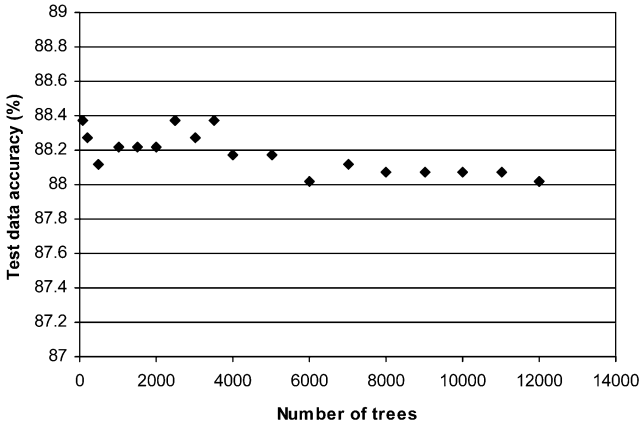

Figure 2. Variation in classification accuracy with a varying number of trees using three features.

kernel function with kernel width $\gamma=2$ and regularization parameter $C=5000$ was used. The 'one against one' method was used to generate a multi-class classifier (Chang and Lin 2001). For further details of SVMs in remote sensing classifications, readers are referred to Pal and Mather (2003b). All the processing with support vector machines was done on a Sun workstation while a window based Pentium IV processor was used for the random forest classifier.

Tables 2 and 3 provide per-class while table 4 provides overall classification accuracies achieved by random forest and SVM classifier, suggesting a comparable performance by both classifiers. The time and experimentation required to select the user-defined parameters for random forest classifier are quite small in comparison to SVMs for this dataset (table 4). This is because the design of SVMs involves choosing a suitable kernel, kernel specific parameters and the regularization parameter C, which requires a lot of experimentation and processing time. Classification accuracy changes from 88.37% to 88.02% when the number of trees increases from 100 to 12000 (figure 2). This indicates that the random forest classifier is almost insensitive to overfitting for this dataset.

Table 2. Confusion matrix by random forest classifier using three features and 100 trees.

Class	1	2	3	4	5	6	7	Total	User's accuracy
1	289	5	2	1	0	0	0	297	97.31
2	6	242	14	8	15	8	2	295	82.03
3	2	15	251	12	2	8	2	292	85.96
4	3	9	24	263	2	5	2	308	85.39
5	0	13	0	2	278	1	0	294	94.56
6	0	15	8	12	3	267	21	326	81.9
7	0	1	1	2	0	11	210	225	94.17
Total	300	300	300	300	300	300	237	2037	
Producer's accuracy	96.33	80.67	83.67	87.67	92.67	89	88.61		

Table 3. Confusion matrix by support vector machines using a 'one against one' multi-class approach.

Class	1	2	3	4	5	6	7	Total	User's accuracy
1	290	5	3	3	0	0	0	301	96.35
2	2	249	9	7	21	6	1	295	84.41
3	6	20	244	10	2	7	1	290	84.14
4	2	7	37	263	1	6	5	321	81.93
5	0	8	0	1	275	0	0	284	96.83
6	0	11	7	16	1	267	27	329	81.16
7	0	0	0	0	0	14	203	217	93.55
Total	300	300	300	300	300	300	237	2037	
Producer's accuracy	96.67	83	81.33	87.67	91.67	89	85.65		

Table 4. Final classification accuracy, kappa coefficient and training time of both classifiers.

Classifier used	Random forest classifier	Support vector machines
Accuracy (%) and kappa value	88.37 (0.864)	87.9 (0.86)
Training time	12.98 s on P-IV	18 s on Sun machine

4. Conclusions

The results reported in §3 suggest that the random forest classifier can achieve a classification accuracy which is comparable to that achieved by SVMs. Another advantage of the random forest classifier is that it requires two parameters only to be set whereas the SVMs require a number of user-defined parameters. The random forest classifier can handle categorical data, unbalanced data as well as the data with missing values, which is still not possible with SVMs. This classifier also provides the relative importance of different features during the classification process, which can be useful in feature selection. Further, the random forest classifier provides a way to detect outliers by using proximity analysis and can be used for unsupervised learning. At present, investigations are still in progress to further evaluate the performance of the random forest classifier for outlier detection, feature selection and clustering with remote sensing data.

Acknowledgment
The author is grateful for the critical comments of two anonymous referees, whose advice has led to an improvement in the presentation of this paper.

References

Boser, B., Guyon, I. and Vapnik, V.N., 1992, A training algorithm for optimal margin classifiers. *Proceedings of the 5th Annual Workshop on Computer Learning Theory* (San Mateo, CA: Morgan Kaufman), pp. 144–152.

Breiman, L., 1996, Bagging predictors. *Machine Learning*, **26**, pp. 123–140.

Breiman, L., 1999, Random forests—random features. Technical Report 567, Statistics Department, University of California, Berkeley, ftp://ftp.stat.berkeley.edu/pub/users/breiman.

Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J., 1984, *Classification and Regression Trees* (Monterey, CA: Wadsworth).

Briem, G.J., Benediktsson, J.A. and Sveinsson, J.R., 2002, Multiple classifiers applied to multisource remote sensing data. *IEEE Transactions on Geoscience and Remote Sensing*, **40**, pp. 2291–2299.

Chang, C. and Lin, C., 2001, LIBSVM: A Library for Support Vector Machines. Department of Computer Science and Information Engineering, National Taiwan University, Taiwan, http://www.csie.ntu.edu.tw/~cjlin/libsvm.

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

Cristianini, N. and Shawe-Taylor, J., 2000, *An Introduction to Support Vector Machines* (Cambridge: Cambridge University Press).

Dietterich, T.G., 2002, *Ensemble Learning. The handbook of brain theory and neural networks*, M.A. Arbib (Ed.) (Cambridge, MA: The MIT Press).

Feller, W., 1968, *An Introduction to Probability Theory and its Application*, vol. 1, 3rd edn (New York: Wiley).

Freund, Y. and Schapire, R., 1996, Experiments with a new boosting algorithm. *Proceedings of the Thirteenth International Conference on Machine Learning*, pp. 148–156.

Friedl, M.A., Brodley, C.E. and Strahler, A.H., 1999, Maximizing land cover classification accuracies produced by decision tree at continental to global scales. *IEEE Transactions on Geoscience and Remote Sensing*, **37**, pp. 969–977.

Giacinto, G. and Roli, F., 1997, Ensembles of neural networks for soft classification of remote sensing images. *Proceedings of the European Symposium on Intelligent Techniques, European Network for Fuzzy Logic and Uncertainty Modelling in Information Technology, Bari, Italy*, pp. 166–170.

Mingers, J., 1989, An empirical comparison of pruning methods for decision tree induction. *Machine Learning*, **4**, pp. 227–243.

Muchoney, D., Borak, J., Chi, H., Friedl, M., Gopal, S., Hodges, J., Morrow, N. and Strahler, A., 2000, Application of MODIS global supervised classification model to vegetation and land cover mapping of Central America. *International Journal of Remote Sensing*, **21**, pp. 1115–1138.

Pal, M. and Mather, P.M., 2003a, An assessment of the effectiveness of decision tree methods for land cover classification. *Remote Sensing of Environment*, **86**, pp. 554–565.

Pal, M. and Mather, P.M., 2003b, Support vector classifiers for land cover classification. Map India 2003, New Delhi, 28–31 January, www.gisdevelopment.net/technology/rs/pdf/23.pdf.

Quinlan, J.R., 1993, *C4.5: Programs for Machine Learning* (San Mateo, CA: Morgan Kaufmann).

Vapnik, V.N., 1995, *The Nature of Statistical Learning Theory* (New York: Springer-Verlag).