World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:1, No:12, 2007

Ensemble Learning with Decision Tree for Remote Sensing Classification

Mahesh Pal

Abstract**—**In recent years, a number of works proposing the combination of multiple classifiers to produce a single classification have been reported in remote sensing literature. 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. As accuracy is the primary concern, much of the research in the field of land cover classification is focused on improving classification accuracy. This study compares the performance of four ensemble approaches (boosting, bagging, DECORATE and random subspace) with a univariate decision tree as base classifier. Two training datasets, one without ant noise and other with 20 percent noise was used to judge the performance of different ensemble approaches. Results with noise free data set suggest an improvement of about 4% in classification accuracy with all ensemble approaches in comparison to the results provided by univariate decision tree classifier. Highest classification accuracy of 87.43% was achieved by boosted decision tree. A comparison of results with noisy data set suggests that bagging, DECORATE and random subspace approaches works well with this data whereas the performance of boosted decision tree degrades and a classification accuracy of 79.7% is achieved which is even lower than that is achieved (i.e. 80.02%) by using unboosted decision tree classifier.

Keywords**—**Ensemble learning, decision tree, remote sensing classification.

Open Science Index, Computer and Information Engineering Vol:1, No:12, 2007 publications.waset.org/717/pdf

I. Introduction

An ensemble is defined as a set of individually trained classifiers whose predictions are combined when classifying a new data. Within last decade several different ways of creating ensemble of classifiers are suggested. In remote sensing [1,2] suggested different way of combining several neural network classifiers, while [3] integrated the classification results of statistical and neural of classifiers. These studies suggest that combined classifiers perform better than the individual classifier used in making ensemble. Other ensemble approaches, which are currently in use, are bagging [4] and boosting [5]. Both of these techniques manipulate the training data to generate multiple classifiers using same base classifier. A number of iterations are carried out with different training data set and results are combined by weighted or unweighted voting to classify unknown examples. Studies carried out using boosting with a univariate decision tree classifier suggest that the resulting classifier perform quite well in comparison with individual classifier [6,7,8] proposed a new technique of creating ensemble, called DECORATE (Diverse Ensemble Creation by Oppositional Relabeling of Artificial Data). This technique works by adding some artificial data to the exiting training data in each iteration of the base classifier while creating ensemble and found to work well in comparison to boosting and bagging [9]. Ho [10] proposed random subspace method of creating ensemble of decision tree utilizes the random selection of attributes or features in creating each decision tree. He used a randomly chosen 50% of the attributes to create each decision tree in an ensemble and the ensemble size was 100 trees.

Hansen and Salamon [11] suggested that combining the output of several classifiers to create an ensemble is useful if these classifiers disagree on some of the input data. This disagreement can be called as the *diversity* of the ensemble. Krog and Vedelsby [12] suggested that diversity in ensemble is an important property of a good ensemble technique. Bagging and boosting creates diversity in an ensemble by sampling and re-weighing the available training data. This paper discusses the results of ensemble classifiers by using bagging, boosting, DECORATE and random subspace with a univariate decision tree as base classifier.

M. Pal is with the National Institute of Technology, Kurukshetra, 136119, Haryana, India (phone: 0091 1744 233356; fax: 0091 1744 238050; e-mail: mpce_pal@yahoo.co.uk).

II. Ensemble Methods

Ensemble combines the output of several classifier produced by the weak learner into a single composite classification. Further, ensemble methods can be used to reduce the error of any weak learning algorithm that consistently generates classifications on various distributions over the training data. Ensemble techniques used in this study i.e. boosting, bagging, DECORATE and random subspace are discussed below.

A. Boosting

Boosting is a method used to improve the accuracy of any classifier by producing a series of classifiers. The training set chosen for a classifier depends on the performance of it earlier classifier. Sample, which is incorrectly classified by an earlier classifier, is selected more often than a correctly classified. Thus, boosting produce a new classifier, which is able to perform well on the new data set. In this study, a boosting algorithm called AdaBoost M1 [5]. Boosting assigns a weight to each observation - the higher the weight; the more that observation influences the classifier. At each trial, the vector of weights is adjusted to reflect the performance of the corresponding classifier, with the result that the weight of misclassified observations is increased. The final classifier aggregates the classifiers generated after each iteration by voting and each classifier's vote is a function of its accuracy.

B. Bagging

Brieman [4] suggest another technique, called bootstrap aggregating or bagging, to improve the accuracy of a base classifier by creating a number of classifiers by manipulating the training data. In this method, each classifier's training set is generated by randomly drawing, with replacement, N examples, where N is the size of the original training set. In this situation, many of the original examples may be repeated in the resulting training set while

Open Science Index, Computer and Information Engineering Vol:1, No:12, 2007 publications.waset.org/717/pdf

others may be left out. The learning system generates a classifier from the sample and aggregates all the classifiers generated from the different trial to form the final classifier. To classify an instance, every classifier records a vote for the class to which it belongs, and the instance is labelled as a member of the class with the most votes. If more than one class jointly receives the maximum number of votes, then the winner is selected using some simple mechanism, e.g., random selection. From the above, it can be seen that bagging is a very simple algorithm.

C. DECORATE

Diverse Ensemble Creation is by Oppositional Relabeling of Artificial Data (DECORATE) approach to create an ensemble work by training a classifier on a given training data set. In successive iterations some artificial data are generated from the training data in a way that their class labels differ maximally from the predicted classes by current ensemble. These data are added to the exiting training data and a new classifier is generated on this new data set. This addition of new data helps to increase the diversity of the ensemble. DECORATE uses mean and standard deviation of the training data and their Gaussian distribution to create a new artificial data set. These artificially generated examples are labelled based on the predications of current ensemble. The number of artificial data is a fraction of the training data which is defined in the start of creating ensemble. Labels are selected in a way that probability of selection of a class is inversely proportional to the prediction of current ensemble. This process also involves in rejecting a new classifier if its addition to the exiting ensemble decreases its accuracy. After each iteration artificial data are removed and process is repeated with a new artificial data every time until the maximum number of iterations are performed.

D. Random Subspace

Random Subspace Method [10] uses a subset of randomly selected features and assigned to an arbitrary learning algorithm. This way, one obtains a random subspace of the original feature space, and constructs classifiers inside this reduced subspace. The aggregation is usually performed using weighted voting on the basis of the base classifiers accuracy. It has been shown that this method is effective for classifiers having a decreasing learning curve constructed on small and critical training sample sizes.

III. Decision Tree Classifier

In the usual approach to classification, a common set of features is used jointly in a single decision step. An alternative approach is to use a multistage or sequential hierarchical decision scheme. The basic idea involved in any multistage approach is to break up a complex decision into a union of several simpler decisions, hoping the final solution obtained in this way would resemble the intended desired solution. Hierarchical classifiers are a special type of multistage classifier that allows rejection of class labels at intermediate stages. Classification trees offer an effective implementation of such hierarchical classifiers. Indeed, classification trees have become increasingly important due to their conceptual simplicity and computational efficiency. A decision tree classifier has a simple form which can be compactly stored and that efficiently classifies new data. Decision tree classifiers can perform automatic feature selection and complexity reduction, and their tree structure provides easily understandable and interpretable information regarding the predictive or generalisation ability of the classification. To construct a classification tree by heuristic approach, it is assumed that a data set consisting of feature vectors and their corresponding class labels are available. The decision tree is then constructed by recursively partitioning a data set into purer, more homogenous subsets on the basis of a set of tests applied to one or more attribute values at each branch or node in the tree. A number of approaches have been developed to split the training data at each internal node of a decision tree into regions that contain examples from just one class, and this is the most important element of a decision tree classifier. These algorithms either minimise the impurity of the training data or maximise the goodness of split. There are many approaches to the selection of attributes used for decision tree induction, and these approaches have been studied in detail by researchers in machine learning [13,14,15,16,17]. The procedure of creating a tree classifier involves three steps: splitting nodes, determining which nodes are terminal nodes, and assigning class label to terminal nodes. The assignment of class labels to terminal nodes is straightforward: labels are assigned based on a majority vote or a weighted vote when it is assumed that certain classes are more likely than others. A tree is composed of a root node (containing all the data), a set of internal nodes (splits), and a set of terminal nodes (leaves). Each node in a decision tree has only one parent node and two or more descendent nodes. A data set is classified by moving down the tree and sequentially subdividing it according to the decision framework defined by the tree until a leaf is reached. Decision tree classifiers divide the training data into subsets, which contain only a single class. The result of this procedure is often a very large and complex tree. In most cases, fitting a decision tree until all leaves contain data for a single class may overfit to the noise in the training data, as the training samples may not be representative of the population they are intended to represent. If the training data contain errors, then overfitting the tree to the data in this manner can lead to poor performance on unseen cases. To reduce this problem, the original tree can be pruned to reduce classification errors when data outside of the training set are to be classified.

IV. Data and Results

Study area used in the reported work is located near the town of Littleport in UK. For the Littleport area, ETM+ data acquired on 19th June 2000 are used. The classification problem involves the identification of seven land cover types (wheat, potato, sugar beet, onion, peas, lettuce and beans). For this study, field data printouts for the relevant crop season were collected from farmers and their representative agencies, and other areas were surveyed on the ground to prepare the ground reference images. A total of 4737 were selected using equalised random sampling plan. To remove any bias in using same pixels for training and testing, total selected pixels were divided in two parts. Out of which 2700 pixels were used for training while remaining 2037 pixels were used for testing both neural and decision tree classifiers.

Classifications were performed in order to evaluate the effects of booting, bagging, DECORATE and random

subspace on the accuracy of the output from decision tree classifier. A total of 50 iterations were used to create ensembles with all four ensemble approaches. While using a decision tree classifier gain ratio as attribute selection measures and error-based pruning was used to prune overgrown decision tree. To study the effect of noise on classification accuracy achieved by different ensemble approaches using a decision tree classifier, a data set having noise level of 20% was created from the training data set of 2700 pixels. The noise was introduced in training data set only. Error was introduced in training data by randomly replacing the class value of some percentage of examples to that of the other values with equal probability. For example, to introduce 20% noise to a training set of 2700 pixels, class values of 540 instances were changed in a way to keep class proportion of each class as it was in original training data.

TABLE I
RESULTS WITHOUT ANY NOISE IN TRAINING DATA

Classifier used	Overall classification accuracy	Kappa value
Decision Tree	83.8	0.811
Decision Tree with boosting	87.43	0.853
Decision Tree with bagging	87.33	0.852
Decision tree with DECORATE	86.74	0.850
Decision tree with random subspace	87.19	0.85

Results from Table I show that all four ensemble approaches perform well with decision tree classifier and an increase of about 3 to 4% in classification accuracy is achieved in comparison to the accuracy achieved with univariate decision tree classifier. The characteristics of training data set used with a classifier have a considerable influence on the accuracy of resulting classification. However, acquiring noise free training data may be difficult and costly for land cover classification problems. Thus, several classifications were carried out to evaluate the effect of noise in training data on the classification accuracy using different ensemble approaches with univariate decision tree classifier. Results suggest that bagging works well with noisy data in comparison to other ensemble approaches with decision tree classifier (Table II). Further, results also suggest that random subspace and DECORATE also works well with noisy data in comparison boosting.

TABLE II
RESULTS WITH 20% NOISE IN TRAINING DATA

Classifier used	Overall classification accuracy	Kappa value
Decision Tree	80.02	0.77
Decision Tree with boosting	79.68	0.76
Decision Tree with bagging	85.47	0.83
Decision tree with DECORATE	83.65	0.81
Decision tree with random subspace	84.88	0.82

V. CONCLUSION

This study compares the performance of four ensemble approaches with decision tree as a base classifier. A major conclusion of this study is that all four ensemble approaches works well with this data and boosting provides slightly better accuracy in comparison other ensemble approaches. Another conclusion from this study is that boosting based ensemble approach is severely affected by the presence of noise in comparison to other ensemble approaches.

REFERENCES

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

[2] J. A.,Benediktsson, J. R.,Sveinsson, O. K.,Ersoy and P. H., Swain, Parallel consensual neural networks. *IEEE Trans. Neural Networks*, 8, 1997, 54–65.

[3] F. Roli, G. Giacinto and G. Vernazza, Comparison and combination of statistical and neural networks algorithms for remote-sensing image classification. Neurocomputation in Remote Sensing Data Analysis, Austin, J., Kanellopoulos, I., Roli, F. and Wilkinson G. (Eds.), Berlin: Springer-Verlag, 117-124, 1997.

[4] L., Breiman, Bagging predictors, *Machine Learning*, 26, 1996, 123-140.

[5] Y.Freund and R. Schapire, Experiments with a new boosting algorithm. *Machine Learning: Proceedings of the Thirteenth International conference*, 148-156, 1996.

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

[7] M. Pal and P. M. Mather, Decision tree classifiers and land use classification. *Proceedings of the 27th Annual Conference of the Remote Sensing Society*, 12-14 September, London, UK, 2001.

[8] G. J., Briem, J. A.,Benediktsson, and J. R., Sveinsson, Multiple Classifiers Applied to Multisource Remote Sensing Data *IEEE Transactions on Geoscience and Remote Sensing*, 40, 2002, 2291-2299.

[9] P. Melville and R. Mooney, Constructing diverse classifier ensembles using artificial training examples. *In Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence*, 505-510, 2003, Acapulco, Mexico, August.

[10] T.K.Ho, The Random Subspace Method for Constructing Decision Forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20, 1998, 832-844.

[11] L. Hansen and P. Salamon, Neural network ensembles, *IEEE Transactions on Pattern recognition and Machine intelligence*, 12, 1990, 993-1001.

[12] A. Krogh and J. Vedelsby, Neural networks ensembles, cross validation and active learning. In D.S. Touretzky, G. Tesauro, and T.K. Leen, editors, *Advances in Neural Information Processing Systems*, volume 7, pages 107–115, 1995, MIT Press, Cambridge, MA.

[13] L.,Breiman, J.H. Friedman, R.A.,Olshen and C.J.,Stone, *Classification and Regression Trees*, Wadsworth, Monterey, CA, 1984.

[14] S. K.Murthy, S. Kasif and S. Salzberg, A system for induction of oblique decision trees. *Journal of Artificial Intelligence Research*, 2, 1994, 1-32.

[15] I. Kononenko and J. S. Hong, Attribute selection for modelling. *Future Generation Computer Systems*, 13, 1997, 181-195.

[16] J. Mingers, An empirical comparison of selection measures for decision tree induction. *Machine Learning*, 3, 1989, 319-342.

[17] J. R. Quinlan, *C4.5: Programs for Machine Learning*. San Mateo: Morgan Kaufmann, San Francisco, 1993.

Open Science Index, Computer and Information Engineering Vol:1, No:12, 2007 publications.waset.org/717/pdf