Schemes for integrating multiple neural networks for Multispectral Data Classification

A.Senthil Kumar[1], S. Chowdhury and K.L. Majumder

Signal and Image Processing Group
Space Applications Centre
Ahmedabad -380015 INDIA.
[1] skumar@ipdpg.gov.in

Abstract

Supervised classification of remotely sensed data using artificial neural networks (ANNs) is largely limited by the number of nodes and architecture of the ANN, given the limited training samples. The best practice is to train a number of networks and choose the one which gives maximum classification accuracy over the data other than the training data set. Recently, consensus schemes using multiple classifiers have been attempted to overcome this problem. In the present study, four recent techniques proposed for combining multi-neural networks are examined for classifying the satellite imagery. Of these, selector net approach proposed by Partridge and Griffith is found to yield the best classification accuracy.

1 Introduction

Until recently, supervised classification of space-borne remotely sensed data has been achieved traditionally with Maximum Likelihood (ML) approach. The main problem with this statistical method is the assumption that the actual probability density function (pdf) of the class in feature space follows a Gaussian. distribution. As pointed out by Atkinson and Tatnall [1], the earth's feature is too complex to get fit into this simple distribution. Also, it is quite likely that a single class can have *multiple* representations in the feature space. For example, the class like urban can contain features not only buildings and roads but also features as grass and water bodies which are, in many a normal classification context, to be dealt with as separate classes other than the urban. Moreover, it is essential that the all classes should have non-singular (invertible) covariance matrices for implementing the ML.

To overcome these limitations, several research workers have explored artificial neural networks (ANNs) for low and high dimensional data classification problems. Even though the ANNs show great promise in high-dimensional data classification cases [2], the advantage is not significant when applied to low-dimensional multispectral data. It is commonly accepted now that the ML still yields good result for some classes for which the ANN shows poor result, and vice versa [3].

The ANN approach is understood to learn the underlying classwise pdf from the data itself and is, by design, a nonparametric approach. This is true provided that there is no limit posed either on available training sample sizes or by computational resources. With the data complexity mentioned above, the classification performance of the ANN is largely decided by the choice of the net size, given the limited training data size practically possible by ground survey. In fact, the ANN of a fairly reasonable size can learn completely training data set, but it cannot guarantee a good generalization performance over data outside the training set [2]. A way to improve the generalization is to train multiple classifiers and employ appropriate consensus schemes to combine their results. Both multiple ANNs [4-8] as well as combine ML/ANN [3,9] have been explored for this purpose.

In the present work, we are concerned with schemes for integrating the multiple ANNs. The motivation behind the work is to study their performance as applied to remotely sensed data. In the following, we describe briefly the four ways of integrating multiple classifiers. The application of these schemes for classifying a Landsat-TM data is described with their results in Sec. 3. Finally, our conclusions are given in Section 4.

2 Schemes for integrating multiple neural networks

An ensemble of independently trained networks can make a collective decision in several ways. These basically differ by way of defining a *confidence measure* to apply the classifier with a least uncertainty for a particular class.

2.1 Plurality voting

The simplest of all collective decision schemes is plurality voting. This method has been successfully employed for handwritten data classification [4,5]. In this method, each individual classifier represents one score that is assigned to one class label. The label, which serves more than half of the total scores, is taken as the final result.

2.2 Trust voting

In this scheme, we select the label of that network which has highest rejection criterion. This is much similar to Gaudail et al. [6], who proposed a confidence measure based upon Euclidean distance of the pixel under test to two most competitive classes for that pixel. We define an equivalent confidence measure

$$C = \frac{d_2 - d_1}{d_2 + d_1} \qquad (1)$$

as applied to the networks. Here d_2 and d_1 are the actual values of the most activated and the second most activated output nodes of each network. In an ideal case, the input vector is identical to one of the class representatives, resulting in values $d_2 = 1.0$ and $d_1 = 0$, and the maximum confidence measure, unity. If the input vectors for which output values to the nearest and next nearest class are equivalent, $d_2 = d_1$, get the minimum confidence value, zero. The confidence measure of other input vectors falls between these extrema.

2.3 Integration with Fuzzy Integral

The fuzzy integral (FI) aims to estimate the maximal grade of agreement between the objective evidence obtained from the classifier for a test pixel and expectations from the same classifier for the given set of classes. We have recently studied the use of FI for integrating multiple networks as well as for integrating ML/ANN. For full details of the FI description, the reader is referred to Kumar et al. [7].

The computation of the FI is as follows: Let $Y = \{ y_1, y_2, ..., y_n \}$ be a finite set of values, and $h : Y \rightarrow [0,1]$ be a function. The fuzzy integral $\boldsymbol{S}$ is evaluated from h and a parameter, the so-called fuzzy measure g, as

$$\boldsymbol{S} = \max_{i=1}^{n} [\, \min \{ h(y_i), g(A_i) \} \,] \qquad (2)$$

where $A_i = \{y_1, y_2, ..., y_i\}$. The fuzzy measures, $g(A_i)$, are obtained using its additive properties in a recursive manner:

$$g(A_i) = g(y_1) = g^1,$$

$$g(A_i) = g^i + g(A_{i-1}) + \lambda g^i g(A_{i-1}),$$

$$\text{for } i = 2,...,n. \qquad (3)$$

The value λ is determined by solving the equation

$$\lambda + 1 = \prod_{i=1}^{n} (1 + \lambda g^i), \qquad (4)$$

where $\lambda \in (-1, +\infty)$, and $\lambda \neq 0$. This is obtained by solving an (n-1)th degree polynomial equation and finding the unique root greater than -1. The fuzzy measures g can thus be fully determined by the so-called density function g^i.

If we let $X = \{x_1, x_2, ..., x_n\}$ as a set of classes and let $Y = \{ y_1, y_2, ..., y_n \}$ be the set of networks, and A be the pixel under consideration to be recognised, then $h_p : Y \rightarrow [0,1]$ represents the partial evaluation of the object of the pixel X for each class x_p. In other words, $h_p(y_i)$ is an indication of how certain we are in the classification of the pixel A to be in class x_p using the network y_i. Corresponding to each y_i, the degree of importance, g^i, of how important y_i is in the recognition of the each class must be given. These densities are generated from classification of the training data.

2.4 Selector net strategy

Partridge and Griffith [8] proposed this approach for integrating the networks. The output nodes of each independently trained network are again retrained along with original teacher vector by an independent net called selector net. The original input vector can also optionally included while training the selector net. The selector net approach improves generalization in *mixed* classes in particular. The output nodes of individual networks take a particular pattern depending on their trained weights. The selector net gives a consensus decision by learning that pattern voted by majority of the networks.

3 Results and Discussions

To study the performance of above schemes, we have used the Landsat-TM sample imagery data set in the image library of the ERDAS™. There is one-to-one ground truth data available for the entire image. For the present study, we have used the multispectral data comprising TM-2, TM-3 and TM-4 bands. Figure 1 shows the 512 × 512 size image of the Lansat-TM3 band multispectral data. The classwise sample size of the image is given in Table 1. For training the classifiers, a sample size of 200 for each class was extracted by picking randomly from the multispectral data. The trained classifiers are then tested with the full image size data. The classified image is then compared with the ground truth data in validating the accuracy of the classification approach.

Five back-propagation networks of three-layer architecture were employed independently to learn the class features. The hidden units of these networks are 4, 7, 12, 17 and 25. Classification performance of these individual networks is given in Table 2. The number of hidden units of these networks is shown in braces. It can be seen that the percentage accuracy (underlined) is moderate for the bare ground and urban classes. These classes are, of course, important from the viewpoint of applications. The fact for the moderate classification is that these classes have common spectral features, resulting in a high confusion. From the table, the average percentages of only these two classes can be shown to be 39.9, 45.17, 44.83, 41.15 and 45.93 respectively.

The classification results obtained with the consensus schemes are given in Table 3. As can be seen, their overall classification accuracy is comparable to those of the individual networks. In particular, the fuzzy integral and selector net methods score over other schemes besides over the individual networks. It is note worthy to compare the average percentage accuracy of the two classes (bare ground and urban) in question. The values are 42.79, 43.81, 48.24 and 53.16 respectively for the plurality, trust voting, FI and the selector net methods.

The poor performance of the voting schemes can be attributed to the fact that there is no control process built in their design. On the other hand, the fuzzy integral, for example, employs the degree of importance of each network to class in question serving a control factor in deciding the collective decision output. Similarly, retraining the output of individual networks makes the selector net learn the complexity of the data; here feeding the teacher signal to the selector net helps to control the decision making. A detailed study on the process of decision making by the consensus schemes is underway.

4 Concluding remarks

The classification performance of integrated neural networks via different schemes were evaluated for multispectral data classification. Each method has given comparable or high classification performance as compared to the individual networks. In particular, the selector net method proves to be the best in terms of overall performance. Work is in progress in order to extend these methods for many other multispectral data sets as well as for integrating data from multiple sources.

5 Acknowledgement

The authors wish to thank Mr. A.K.S.Gopalan, Director, SAC, for constant encouragement during the course of study.

References

[1] P.M. Atkinson and A.R.L. Tatnall, “Neural networks in remote sensing : An Introduction,” *Int. J. Remote Sensing*, vol. 18, 699-709, 1997.

[2] J. A. Benediktsson, P.H. Swain, and O.K. Ersoy, “Conjugate-gradient neural networks in classification of multisource and very high-dimensional remote sensing data,” *Int. J. Remote Sensing*, vol.14, 2883-2903, 1993.

[3] A. Pinz and P. Bartl, “Information fusion in image understanding: Landsat classification and ocular fundus images” in *Sensor Fusion-V*. The SPIE, vol. 1828, 276-287, 1992.

[4] L.K. Hansen and P. Salamon, "Neural network ensembles", *IEEE Trans. PAMI,* vol.12, 993-1001, 1990.

[5] L. Xu et al.," Methods of combining multiple classifiers and their applications to

handwriting recognition," *IEEE Trans. SMC,* vol.22, 418-434, 1992.

[6] F. Goudail, E.Lange, T. Iwamoto, K. Kyuma and N.Otsu, "Face recognition system using local autocorrelations and multiscale integration", *IEEE Trans. PAMI,* vol.18, 1024-1028, 1996.

[7] A. S. Kumar, S.K. Basu and K.L. Majumder, “Robust classification of multispectral data using multiple neural networks and fuzzy integral” *IEEE Trans. on Geoscience and Remote Sensing*, vol. **35**(3), 787-790, 1997.

[8] D. Partridge and N. Griffith, "Strategies for improving Neural Network Generalizaion", *Neural Comput. Applications*, vol.3 27-37, 1995.

[9] A. S. Kumar and K.L. Majumder, “Information fusion in tree classifiers”, *Int. J. Remote Sensing* (in press). ; see also, A.S. Kumar, S. Chowdhury and K.L.Majumder, "Combination of neural and statistical approaches for classifying space-borne multispectral data", in: *Int. Conference on Advances in Pattern Recognition and Digital Techniques,* Ind. Stat. Inst., Calcutta, Dec. 27-29, pp.87-91, 1999.

Table 1. Classes and their sample sizes in test image.

Classes	No. of samples
Forest	176987
Water	23070
Agriculture	26986
Bare ground	740
Grass	12518
Urban	11636
Shadow	3197
Clouds	350
Total Samples	262144

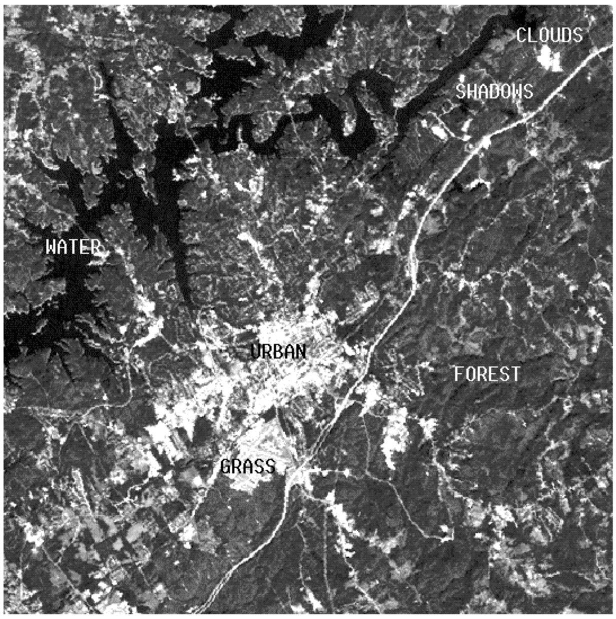

Figure 1. One of multispectral images (LANDSAT-TM3) used in the study.

Table 2. Classification results with independently trained networks

Classes	NN(4)	NN(7)	NN(12)	NN(17)	NN(25)
Forest	74.33	75.46	75.51	75.72	75.53
Water	87.58	85.79	86.88	86.13	87.41
Agriculture	86.80	89.18	86.06	87.60	88.26
Bare ground	56.80	46.49	39.14	41.91	40.30
Urban	23.01	43.85	50.52	40.38	51.56
Grass	75.19	78.05	84.13	83.52	81.86
Shadow	92.90	96.43	96.81	97.84	97.00
Clouds	92.86	93.43	93.43	94.57	91.71
Average	73.68	76.08	76.56	75.96	76.70

Table 3. Classification results with consensus schemes

Classes	Plurality	Trust voting	Fuzzy integral	Selector net
Forest	75.35	74.94	75.88	75.51
Water	87.15	87.36	87.08	91.89
Agricultur	87.73	87.77	85.84	82.34

e				
Bare ground	42.38	42.01	46.30	58.11
Urban	43.19	45.60	50.17	48.21
Grass	82.46	83.30	79.39	77.73
Shadow	97.84	98.00	97.72	94.28
Clouds	93.14	93.71	92.29	92.57
Average	76.16	76.59	76.83	77.58