A Combination of Temporal Thresholding and Neural Network Methods for Classifying Multiscale Remotely-Sensed Image Data

A. Moody, S. Gopal, A. H. Strahler, J. Borak and P. Fisher

Department of Geography, Center for Remote Sensing
Boston University
675 Commonwealth Avenue
Boston, MA 02215 (USA).
T: 617.353.5981 EMail: moody@crsa.bu.edu

ABSTRACT

A thresholding technique applied to a time series of 1 km, biweekly composited AVHRR-NDVI data stratifies a Sierra Nevada test site into broad vegetation classes based on temporal habit. These classes provide a basis for developing strata specific neural network classifiers which operate on single date 250 m and 500 m data simulated from Landsat Thematic Mapper, as well as 1000 m digital elevation data. The results of this combined hierarchical approach are compared to results from a network developed for the site as a whole without prior stratification. The artificial neural networks are feedforward models based on the multilayer perceptron structure trained by a backpropagation algorithm. The combined approach appears to perform slightly better than the sitewide model, although the results are comparable. The results from both approaches produced cover-type proportions which are closer to the proportions in the original 30 m reference class map than to the aggregated 250 m map used for training and testing the network models.

INTRODUCTION

The application of artificial neural networks to the extraction of information from remotely sensed data is attracting increasing interest from numerous scientific perspectives. In particular, neural networks have been applied to remote sensing classification tasks with notable successes (Benediktsson et al., 1990; Kanellopoulos et al., 1992). This has led to the consideration of using neural networks as part of the global land-cover classification algorithm currently being developed for processing data from the planned MODIS instrument (Moderate Resolution Imaging Spectroradiometer) due for launch aboard the EOS-AM platform in 1998 (Moody et al., 1994). MODIS will provide 2-day coverage of the globe and will produce red and near infrared data at 250 m resolution, visible and near infrared data at 500 m and middle and thermal infrared data at 1 km. Other possible MODIS land-cover algorithms will rely on temporal-threshold based decision rules to stratify land cover into behavioral/structural forms (Running et al., 1994). This paper represents a preliminary investigation into the integration of these two techniques for a Northern California, Sierra Nevada test site. The basic premise is that thresholding algorithms applied to time trajectories of the coarser scale data can provide an initial stratification of cover types, and single-date or reduced temporal frequency, high resolution data can further refine the classification based on strata specific neural network algorithms. This speaks in part to the definition and development of algorithms for global land-cover characterization from MODIS, as well as to the understanding of how the combination of spatial resolutions that MODIS will provide can best be employed. The purpose of this work is to compare the classification performance of the integrated two level processing approach to a neural network classifier developed for the test site as a whole.

TEST DATASET

This research employs data from the Plumas National Forest in California. The Plumas is roughly 8000 sq. km. and lies at the transition between the northern extent of the Sierra Nevada Mountains and the southern edge of the Cascade Range. This site has been studied recently as part of a project to develop land-cover mapping and timber-inventory methods for the U. S. Forest Service (Woodcock et al., 1993). A vegetation map was produced using Landsat Thematic Mapper imagery and unsupervised image classification supported by air-photo and field validation. This map has been aggregated to 250 m and serves as a reference for the work presented here. Aggregation was performed using a simple plurality rule over a 240 m grid and the resulting map was then resampled to 250 m. Small random samples of "pure" class pixels were extracted from this data for training of the networks. Cover classes for this dataset include *grass/barren*, *brush*, *hardwood*, *meadow*, *conifer* and *water*. Meadows are omitted from this study due to their small size and relative infrequency. Inputs are June 1990 TM data which have been spatially filtered to simulate MODIS 250 m data in red and near infrared bands 3 and 4, and 500 m data in bands 1, 2, 5, and 7. Biweekly composited 1 km AVHRR-NDVI data from March to December 1990 is coregistered to the filtered TM data and the resampled vegetation map. A digital elevation model at 1 km resolution is also registered and used as input.

THRESHOLD METHOD

The temporal thresholding technique involved a series of decision rules applied to AVHRR-LAC data. This is a largely modified version of a similar approach suggested by Running et al. (1994). These data were in the form of biweekly composited NDVI images. The complete data set for the procedure consisted of nineteen such images, representing nearly complete temporal coverage for 1990. The first decision involved selection of representative images for the growing and non-growing seasons. The purposes of this procedure were two-fold. During this phase significantly snow-covered images were removed from the non-growing season data. Also, late spring and early summer dates were selected as representative of the growing season in order to avoid the use of data collected during periods of strong seasonal transitions.

After selecting the representative images, an NDVI threshold of 0.35 was applied on a per-pixel basis to the entire representative subset in order to discriminate between vegetated and non-vegetated surface features. Where the values of particular pixels failed to exceed this threshold in every image of the representative subset, the pixels were designated barren of vegetation. The next decision concerned discrimination of annual and perennial vegetation types. Where the values of particular pixel types exceeded the barren threshold (NDVI=0.35) during the growing season, but

0-7803-1497-2/94 $4.00 © 1994 IEEE

failed to do so during the non-growing season, the pixels were labeled annual vegetation. Those pixels that exceeded this threshold over both seasons were designated as perennial vegetative features.

The final step in the thresholding procedure further divided the perennial vegetation category into deciduous and evergreen classes. For each pixel associated with perennial vegetation, mean NDVI values were calculated for both growing and non-growing seasons. These means were compared on a per-pixel basis via a Student's t-test for samples of unequal variance. Pixels exhibiting statistically significant differences in NDVI across the seasons were labeled deciduous, while those pixels that were not statistically different were designated evergreen. Using this technique, the test site was stratified into *annual, barren, deciduous* and *evergreen* strata.

NEURAL NETWORK METHOD

This research employs fully connected, feedforward network models which are based on the multilayer perceptron structure and trained by the backpropagation algorithm (Rumelhart and McClelland, 1986). The network is composed of layers of "neurons" which are interconnected through weighted synapses. All nodes in a given layer are connected to each node in the subsequent layer of the network and each connection has an associated weight which can be excitatory or inhibitory. The first layer represents the classification input variables and the last layer represents the output classes. Intermediate "hidden" layers represent an internal representation or neural pathways through which input data are processed to arrive at output values or conclusions. Through an iterative presentation of input/output pairs, the network can "learn" to recognize patterns of input signals and relate them to desired output responses. In a supervised approach input patterns are first fed forward through the network. The internal structure of the network, or hidden layers, allow interactions between inputs to develop and provide the basis for discrimination surfaces. Outputs are linear combinations of throughput signals and the synapse weights. During the learning phase, errors are calculated as the RMS error between the network outputs and the desired outputs and are backpropagated through the network. On each iteration, the synapse weights are adjusted in order to reduce the total RMS error until a convergence criterion has been satisfied. Once the network has been trained, it can be applied to new input data and evaluated.

For this analysis, all networks have seven inputs (spatially filtered TM bands 1, 2, 3, 4, 5, and 7 and the DEM data) and six outputs (*grass/barren, brush, hardwood, water, conifer* and *none*). Each network has only one hidden layer, but the number of hidden nodes varies between the models for the different strata. The different models have varying momentum terms, learning rates, convergence criteria and and varying numbers of training samples (Table 1). All of the networks employed the delta learning rule and a sigmoid transfer function. Network training was based on small subsets of relatively "pure" pixels selected from each class (see Table 1). Roughly 2.5% (5000 pixels) of the total unmasked image area was used for training. For the combined approach, networks were developed separately for each strata defined using the temporal thresholding method. The results of this method are compared to a network which was developed for the entire test site without prior stratification.

ANALYSIS OF RESULTS

Both the combined threshold/network approach and the sitewide network model were trained and applied to the Plumas test site and results were evaluated using the aggregated 250 m class map. Tables 2 and 3 are confusion matrices showing results from the two methods. The first value at each matrix element represents the number of pixels falling into that confusion category. The second element is the percent confusion with respect to the size of the reference map class, and the third element (in parentheses) is the percent confusion with respect to the size of the output map class.

The combined approach has better correct classification performance for *brush, hardwood* and *water* when evaluated with respect to the reference map. For example, for the pixels which were actually classified as *hardwood* on the reference map, 76% were also classified as *hardwood* by the combined approach and 73% were classified as *hardwood* by the sitewide network. With respect to the network outputs, the combined approach performs better for *grass, water* and *conifer*. For example, of all the pixels which were classified as *conifer* by the combined approach, 87% actually were *conifer* on the reference map. This value is 85% for the sitewide model. On the whole, it appears that the combined approach performs better than the sitewide approach. This is only a slight advantage, with improvements in the range of 2% to 4%, is dependent on the method of evaluation, and is not consistent for all classes. It is interesting to note, however that the combined approach performed as well or better than the sitewide approach with a very small number of within strata training samples (Table 1). For example, the *barren* and *annual* strata employed only 75

Table 1. *Number of hidden units, momentum terms, number of training samples and convergence criteria for the sitewide network and the four strata-specific networks.*

Network Information					
	Threshold Based Strata				
	Sitewide	Barren	Ann.	Decid.	Evrgrn.
Hidden	15	5	8	10	12
Mom.	0.8	0.5	0.8	0.8	0.9
Samples	5000	75	325	1280	3125
Conv.	0.04	0.04	0.04	0.05	0.04

Table 2. *Accuracies for combined method: First entries are pixel quantities, second entries are confusion proportions with respect to reference map, and entries inside parentheses are proportions with respect to output class map. Total number of pixels in each output class are indicated below the class type in the left most column.*

Confusion Matrix for Combined Approach					
Outputs	Reference Map Classes				
	Grss	Brsh	Hrdwd	Wtr	Confr
Grss	7789	5617	63	19	459
13947	0.625	0.229	0.005	0.017	0.007
	(0.56)	(0.40)	(0.005)	(0.001)	(0.03)
Brsh	3529	13780	1260	28	12254
30851	0.283	0.561	0.091	0.024	0.197
	(0.11)	(0.45)	(0.04)	(0.001)	(0.40)
Hrdwd	569	1664	10510	17	8380
21140	0.046	0.068	0.759	0.015	0.135
	(0.03)	(0.08)	(0.50)	(0.001)	(0.40)
Wtr	36	53	100	978	1995
3162	0.003	0.002	0.007	0.855	0.032
	(0.01)	(0.02)	(0.03)	(0.31)	(0.63)
Confr	541	3449	1918	98	39083
45089	0.043	0.140	0.138	0.086	0.628
	(0.01)	(0.08)	(0.04)	(0.002)	(0.87)
Ref Tot	12464	24563	13851	1140	62171

Table 3. *Accuracies for single network method: Entry explanations are the same as in Table 2.*

Confusion Matrix for Sitewide Network					
Output	Reference Map Classes				
	Grss	Brsh	Hrdwd	Wtr	Confr
Grss 15412	8422 0.669 (0.55)	6264 0.254 (0.41)	75 0.005 (0.005)	24 0.021 (0.002)	627 0.010 (0.04)
Brsh 27743	3328 0.265 (0.12)	12837 0.521 (0.46)	1227 0.088 (0.04)	22 0.019 (0.001)	10329 0.166 (0.37)
Hrdwd 18413	127 0.010 (0.01)	1569 0.064 (0.09)	10103 0.727 (0.55)	11 0.010 (0.001)	6603 0.106 (0.36)
Wtr 4639	31 0.002 (0.01)	44 0.002 (0.01)	191 0.137 (0.04)	963 0.833 (0.21)	3410 0.055 (0.74)
Confr 48365	672 0.053 (0.01)	3958 0.160 (0.08)	2308 0.094 (0.05)	136 0.118 (0.003)	41301 0.663 (0.85)
Ref.Tot	12569	24630	13879	1152	62142

and 325 training samples and made up 7% and 11% of the image area, respectively. The networks for these strata therefore trained on only 1.5% and 6.5% of the total training dataset. It is likely that the combined method could be enhanced significantly with a more statistically representative training dataset.

With respect to the 250 m reference class map, both classification approaches overestimated the occurrence of all classes except for *conifer*, which was underestimated. One of the more interesting results of this analysis is that the classification outputs produced overall class proportions which were much closer to the proportions in the original 30 m class map than to the proportions in the 250 m aggregated reference map (Table 4). It is unclear at this point whether this translates into a better locational accuracy for these methods if they were evaluated using the 30 m map. It remains to be tested whether this result is merely coincidental or has a physical explanation that will hold across landscapes. It is likely that the networks are responding to the spectral mixing of subpixel components and/or, that land cover proportions as determined from remotely sensed data may scale more effectively than expected or than represented by the class map aggregation procedure used. If this is the case, then the relatively poor performance of the classifiers may have more to do with distortions in the reference map than with limitations in the information content of the data or the strength of the techniques.

Table 4. *Cover-type proportions for the original 30 m reference map, the 250 m aggregated reference map used for training and testing the models, the combined threshold/network results, and the sitewide network approach.*

Comparison of Cover Proportions				
Classes	Class Maps			
	30m Ref	250m Ref	Combined	Sitewide
Grss	13.91	10.97	12.25	13.75
Brsh	23.54	21.52	27.26	24.72
Hrdwd	13.27	12.14	18.53	15.98
Wtr	0.91	1.01	2.79	4.01
Confr	46.22	54.37	39.10	41.53

CONCLUSION

While both methods tested provide similar results, it is likely that the combined method using preclassification stratification based on temporal thresholding would improve if the networks for the individual strata were trained on more representative samples. While the two stage approach is considerably more labor intensive, it provides an avenue through which multiresolution data can be combined, and through which both high temporal and high spatial frequency data can be exploited to develop large area land-cover datasets. This approach also provides a structure which allows separate networks to be trained and applied over reasonably self-similar strata. On a continental or global scale, some such stratification is needed to allow for regional extensibility of the network algorithms.

The apparent ability of these methods to output cover-type proportions which are close to the high resolution proportions of the scene components needs to be further investigated. This result may point out the need to develop improved methods for scaling categorical spatial data and may also highlight the value of investigating the sensitivity of this type of neural network algorithm to subpixel mixtures of scene components.

Acknowledgements

This work was supported by the National Aeronautics and Space Administration under contract NAS5-31369. The authors gratefully acknowledge John Collins, Vida Jakabhazy, Scott Macomber and Soren Ryherd for their work in producing the Plumas land-cover classification.

REFERENCES

Benediktsson, J.A., Swain, P.H. and Ersoy, O.K. "Neural network approaches versus statistical methods in classification of multisource remote sensing data." *IEEE Trans. Geosci. Rem. Sens.*, 28(1990):540-552.

Kanellopoulos, I., Varfis, A., Wilkinson, G.G. and Megier, J. "Land-cover discrimination in SPOT HRV imagery using an artificial neural network - a 20-class experiment." *Int. J. Remote Sensing*, 13(1992):917-924.

Moody, A., Strahler, A., Huete, A., Justice, C., Muller, J., Running, S., Salomonson, V., Vanderbilt, V. and Wan, Z. *MODIS Land Cover Product Algorithm Technical Basis Document.* NASA document for MODIS Product No. 11, Parameter Nos. 2669 and 2671 (1994).

Rumelhart, D.E., McClelland, J.L. and PDP Research Group (Eds.). *Parallel Distributed Processing: Exploration in the Microstructure of Cognition. Vol. 1: Foundations.* Cambridge MIT Press, Cambridge. 1986.

Running, S.W., Loveland, T.R. and Pierce, L.L. "A remote sensing based vegetation classification for global land cover analysis." *Ambio*, 23 (1994):4-8.

Woodcock, C.E., Collins, J., Gopal, S., Jakabhazy, V., X. Li, X., Macomber, S., Ryherd, S., Wu, Y., Harward, V., Levitan, J. and Warbington, R. "Mapping forest vegetation using Landsat TM imagery and a canopy reflectance model." submitted to *Remote Sens. Environ.* (1993).