CLASSIFYING DRY SCLEROPHYLL FOREST FROM AUGMENTED SATELLITE DATA: COMPARING NEURAL NETWORK, DECISION TREE & MAXIMUM LIKELIHOOD

L.K. Milne[1], **T.D. Gedeon**[1] and **A.K. Skidmore**[2]
[1]School of Computer Science and Engineering
[2]School of Geography
The University of New South Wales

Abstract

Detailed maps derived from geographical data are becoming increasingly desirable for use in forest management. Many types of data are available for use in generating maps, for example, soil and vegetation maps.

We look at a method for giving high level classifications that can be used as additional data for the generation of more detailed maps, and compare the results with other currently used techniques. We use multiple techniques to increase the reliability and accuracy of predictions.

We describe a simple method of adjusting the balance of false positive and false negative classifications that are produced by the neural network. This allows better integration with non-neural network techniques.

Introduction

Collecting data for detailed maps from ground surveys over large areas is prohibitively expensive. The areas being studied can also change considerably over short periods of time, so it is desirable to be able to generate maps of forest attributes automatically.

Satellite data has been used quite successfully to distinguish gross features, such as land and water (Omatu and Yoshida, 1991). It has been found that using satellite data alone does not provide sufficient information for more detailed mapping to be done, such as distinguishing forest species (Skidmore and Turner, 1988).

Ancillary data, such as aerial photographs, have been used to augment the satellite data to achieve better results (Skidmore et al, 1994).

The Data

We are using geographical data which is from an area in the Nullica State Forest on the south coast of New South Wales. The area, approximately 20 by 10 km, is broken up into a grid of 179831 pixels, 30 by 30 m in size.

The data has been collected from satellite imagery, soil maps and aerial photographs. From the aerial photographs it is possible to derive a terrain model and from this derive a number of terrain features. In this case each pixel has a value for altitude, aspect, slope, geology, topographic position, rainfall, temperature, and Landsat TM bands 1 to 7.

For the purpose of training 190 detailed sample plots have been surveyed. This data gives us classifications for 190 of the pixels in the field area. For use in testing we have 70 pixels that have been surveyed in less detail, for which we know the classes.

The data has been preprocessed using a cumulative histogram enhancement technique (Richards, 1986). For each feature, the cumulative histogram gives the total number of pixels that are less than or equal to a given value, for all possible values. The spread of the data in the histogram is smoothed out using the transformation:

$$v' = b_v * (l - 1) / n$$

where b_v is the cumulative histogram bin count for a given value v,

l is the number of values required,
n is the number of pixels, and
v' is the value that v is mapped to.

Using this algorithm the values were scaled to between 0 and 99 and then divided by 100.

The requirement for this preprocessing is demonstrated by the distribution of values for the geology field, in Figure 1.

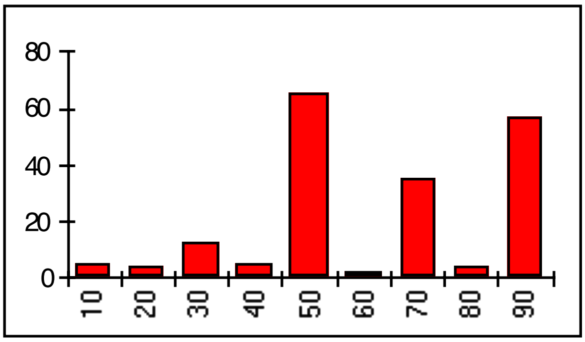

Figure 1. Geology descriptor distribution

As the aspect values are represented in degrees we have a problem of continuity. An aspect of 0 is north, but so is 360. To overcome this, the aspect is represented using vectors such that each vector is a fixed distance from its neighbouring values (Bustos and Gedeon, 1995). That is, the aspect is encoded in four inputs representing the major compass points, with some smaller input values for the adjacent directions. In this way the distance in the input space between North to NE is the same as to NW and so on.

There are five classes for each of the pixels - scrub, dry sclerophyll, wet/dry sclerophyll, wet sclerophyll and rainforest. Initially, we wish to be able to distinguish pixels that contain only dry sclerophyll forest. Of the 190 training pixels, 99 are *Dry*.

Decision Tree Classification

To derive a set of rules for classifying the pixels the C4.5 program was used (Quinlan, 1993). This is a program based on the ID3 algorithm for generating a knowledge base from a data set.

The default class for pixels not covered by the rules above is Dry. The performance of the rules on the training and test sets is shown below:

Note that 109 out of 190 patterns corresponds to 57.4% performance on the training set, while 46 out of 70 patterns is 65.7% performance on the test set.

It is relatively unusual to achieve a higher value on the test set. This may be due to a higher proportion of noisy patterns in the training set. That is, the test set is from the same population as the training set, but is cleaner.

C4.5 performance			
	correct	false +ve	false -ve
training	109	81	0
test	46	24	0

Maximum Likelihood Classification

Maximum likelihood classification is a statistically based method (Richards, 1993). For each pixel, x, the discriminant function $g_i(x)$ is calculated.

$$g_i(x) = -\ln|\Sigma_i| - (x - m_i)^t . \Sigma_i^{-1} (x - m_i)$$

The class i given to a pixel is that with the maximum value that is over the threshold:

$$T_i = -4.774 - \tfrac{1}{2} \ln|\Sigma_i| + \ln p(w_i)$$

where m_i is mean of class i

Σ_i is covariance matrix for class i

The performance of the maximum likelihood classification on the training and test sets is shown below:

Maximum likelihood performance			
	correct	false +ve	false -ve
training	124	24	42
test	42	14	14

The performance is 65.3% on the training set, and only 60% on the test set. This result seems to contradict our possible explanation for the difference in the results between training and test sets using the C4.5 program.

Neural Network Classification

A number of network topologies were tested by varying the numbers of hidden nodes and hidden layers. There was no significant differences in the classifcation abilites of the topologies. The network we have used is a 3 layer network, with a 17 node input layer, a single 14 node hidden layer and a single node output layer. The 17 inputs are 4 aspect values, altitude, slope, geology, topographic position, rainfall, temperature, and TM bands 1 to 7. We have used a standard back-propagation model, and will subsequently just use the words neural network.

The neural network was trained on the training set, until the error on the test set was minimum, to avoid overtraining on the training set. The point at which to stop training was confirmed by a cross-validation performed on the training set.

Kogan (1991) has shown that a neural network trained for categorisation such as ours, can not be used for scoring. That is, we can not use the value of the output node as a confidence factor. If we wished to do the latter, we would need to train networks explicitly using the probabilities of class membership.

One subtle consequence of Kogan's work which we have not seen elsewhere is that the value of 0.5 as a threshold to distinguish between two classes is not sacrosanct either. Thus, in this experiment we vary the threshold θ from 0.4 to 0.75. The results on the training set are shown below, left.

Neural net performance on the training set			
θ	correct	false +ve	false -ve
0.40	99	91	0
0.45	100	90	0
0.50	**100**	**90**	**0**
0.55	107	81	2
0.60	113	47	30
0.65	89	17	84
0.70	**90**	**1**	**99**

The test set results are shown above, right.

The results for the standard threshold of 0.5 are similar to the C4.5 result, with performance of 52.6% on the training set, and a performance of 65.7% on the test set. The network seems similarly to have chosen the *Dry* category as default.

We have varied the threshold θ, between 0.4 and 0.75.

With a threshold of 0.7, the network performance on the training set has dropped slightly, and dropped noticeably on the test set. Note the situation with respect to false positives/negatives has essentially reversed. Where at a threshold of 0.5, there were no false negatives, at a threshold of 0.70 there are no false positives. Note that the threshold of 0.6 maximises performance on the training set without degrading the performance on the test set, with quite roughly similar numbers of false positives and negatives.

For our purposes, in combining the evidence from multiple sources and to thus arrive at better classifications of the forest supratype, we are more interested in producing low numbers in the false positive category. At the threshold of 0.7, the performance on the test set is only 34.3%, but all errors are false negatives. That is, any pixel categorised as *Dry* is a member of the dry sclerophyll forest supratype. Of course, the pixels classified as *not Dry* will include some pixels which are incorrectly labelled. Using the two threshold values of 0.5 and 0.7 allow us to classify some pixels accurately. Intuitively, what we are doing is finding the boundaries of the overlap regions of the classification.

Neural net performance on the test set			
θ	correct	false +ve	false -ve
0.40	46	24	0
0.45	46	24	0
0.50	**46**	**24**	**0**
0.55	46	23	1
0.60	46	16	8
0.65	37	3	30
0.70	**24**	**0**	**46**

We have found these boundaries by looking at the number of occurrences using the test set, moving and stopping moving the threshold in a fashion analogous to the use of test sets and their error values as a validation test to terminate training. Clearly, by setting the threshold to 0.0 or 1.0, we can trivially maximise or minimise the values of the false positive or negative categories. The issue here is to find the boundaries which do not reduce the correct values too much.

Conclusions and Further Work

We have used satellite imaged data of a NSW state forest augmented by ancillary data derived from aerial photography and other available information. The data has been used to derive classifications for dry sclerophyll forest supratype, using maximum likelihood, C4.5, and neural network techniques.

Statistically there is no significant difference in the classification abilities of the three methods. The maximum likelihood classifier produced the worst results on the test data. Both C4.5 and the neural network produced similar results, with some improvement over the maximum likelihood, but both producing overestimations of the number of *Dry* classifications. That is, both produced high false positive results (and low false negative).

We showed how a simple technique of modifying the threshold in accordance with the results on the validation test set, can be used to modify the neural network result to minimise the false positive results while keeping correct results as high as possible.

The next stage in our work will be to use the remaining values in the existing C4.5 and maximum likelihood classifiers, and to retrain both of these as well as a separate neural network to classify those pixels. This hybrid approach shows considerable improvement in identifying areas of dry sclerophyll forest.

References

Bustos, RA and Gedeon, TD "Decrypting Neural Network Data: A GIS Case Study," *Proceedings International Conf. on Artificial Neural Networks and Genetic Algorithms (ICANNGA)*, Alès, 1995.

Kogan, "Neural networks trained for classification can not be used for scoring," *IEEE Trans. on Neural Networks*, 1991.

Omatu, S and Yoshida, T "Pattern Classification for Remote Sensing using Neural Network, *Proceedings International Joint Conference on Neural Networks*, pp. 653-658, Singapore, 1991.

Quinlan, JR *C4.5: Programs for Machine Learning*, Morgan & Kaufmann, 1993.

Richards, J *Remote Sensing Digital Image analysis*, Springer Verlag, 2nd ed., 1993.

Skidmore, AK, Brinkhof, W and Delaney, J "Using Neural Networks to Analyse Spatial Data," *Proceedings 7th Australasian Remote Sensing Conference*, Melbourne, pp. 235-246, 1994.

Skidmore, A and Turner, BJ "Forest Mapping Accuracies are improved using supervised non-parametric classifier with spot data," *Photogrammatic Engineering and Remote Sensing*, vol. 54, no. 10, pp. 1415-1421, 1988.