N93-25976

CLASSIFYING MULTISPECTRAL DATA BY NEURAL NETWORKS

Brian A. Telfer, Harold H. Szu
Naval Surface Warfare Center, Code R44
Silver Spring, MD 20903
btelfe@ulysses.nswc.navy.mil

Richard K. Kiang
Earth Science Data Operations Facility
NASA Goddard Space Flight Center
Greenbelt MD 20771
rkiang@favente.gsfc.nasa.gov

ABSTRACT

Several energy functions for synthesizing neural networks are tested on 2-D synthetic data and on Landsat-4 Thematic Mapper data. These new energy functions, designed specifically for minimizing misclassification error, in some cases yield significant improvements in classification accuracy over the standard least mean squares energy function. In addition to operating on networks with one output unit per class, a new energy function is tested for binary encoded outputs, which result in smaller network sizes. The Thematic Mapper data (four bands were used) is classified on a single pixel basis, to provide a starting benchmark against which further improvements will be measured. Improvements are underway to make use of both subpixel and superpixel (i.e. contextual or neighborhood) information in the processing. For single pixel classification, the best neural network result is 78.7%, compared with 71.7% for a classical nearest neighbor classifier. The 78.7% result also improves on several earlier neural network results on this data.

INTRODUCTION

In the past several years, a general awareness of the environmental crises has gradually taken place among the world's nations. We wish to address automated surveillance technology for environmental issues. Global warming, ozone depletion, large-scale deforestation, extinction of species are just a few of the issues that could lead to serious consequences to all inhabitants on the Earth, in a scale that will respect no national or political boundaries. To understand and quantify the anthropogenic impact on the environment, and to predict the eventualities if the deteriorating trend is not reverted, consistent and long-term monitoring of the global environment is needed. Through the Earth Probes and the Earth Observation System (EOS), NASA's Mission to the Planet Earth will continue to provide the essential measurements.

The amount of measurements from the Mission to the Plant Earth, however, will be unprecedented. For example, the first EOS AM platform alone will generate more than one terabyte (TB) data a day, compared with the 5 TB from the entire 12 years of AVHRR Pathfinder data. To timely process, analyze, store, and disseminate the satellite measurements and extracted information to a worldwide user community presents a formidable challenge, and demands innovative analytical methods and advanced computing and data communication technologies.

Among the contemporary information sciences, neural networks have proven to be a versatile technique for input-to-output mapping, without the constraint of formulating the exact relationship between the two. In addition, contextual and neighborhood knowledge can be easily included. In the past few years, neural networks have been applied to classifications of remotely sensed data (e.g., Campbell et al. 1989, Decatur 1989, Benediktsson et al. 1990, Liu et al. 1991, Bischof et al. 1992, Kiang 1992). In these studies, spectral data and ground truth are input to multilayer perceptron networks with one or more hidden layers, and networks are extensively trained off-line by minimizing a least-mean-squares (LMS) energy function with back-propagation (Werbos 1974, Rumelhart et al. 1986). It has been shown that the performance of neural network techniques is superior to classical techniques for systems operating in real-time.

It is well documented that minimizing the LMS energy function produces a neural network that approximates the Bayesian *a posteriori* probabilities (the

probability of a class given a particular input vector) of classes of data represented by a training set (see Richard & Lippmann 1991, for a review). Given an infinitely large training set and a network with sufficient functional complexity, the approximation error becomes negligible, and the misclassification error converges to the Bayes rate. While this property makes the LMS energy function attractive, there is an important qualification. The functional complexity needed for approximating the *a posteriori* probabilities is greater than that needed for approximating only class boundaries. Thus, if we are only interested in the classification of an input, rather than its *a posteriori* probability, a neural network that estimates probabilities will be needlessly complex. The additional complexity is a disadvantage both from the principle of parsimony (using the smallest number of weight parameters to increase generalization [see, e.g., Barron & Cover 1991]) and from the hardware implementation standpoint. Therefore, we test energy functions that minimize the misclassification error directly (Szu & Telfer 1991, Telfer & Szu 1992a), rather than indirectly via approximating the *a posteriori* probabilities. We call these Minimum Misclassification Error (MME) energy functions.

We first formulate these energy functions and provide a two-feature example that illustrates the concept. The Landsat Thematic Mapper data is described and results are presented for classifying on a pixel-by-pixel basis. These results are intended to provide a benchmark for further improvements that make use of both subpixel and superpixel (contextual) information. The paper concludes by discussing these research directions.

ENERGY FUNCTION FORMULATION

The commonly used σ-LMS energy function is given by

$$E_{\sigma-LMS} = \sum_{n=1}^{N}\sum_{k=1}^{K}[d_{nk} - \sigma(o_{nk})]^2, \qquad (1)$$

where d_{nk} is the desired output (normally set to 0 or 1) of the k-th output unit for the n-th training vector, σ is a sigmoidal function [we use $\sigma(z) = 1/(1+exp(-z))$], and o_{nk} is the output of the k-th output unit for the n-th training vector, before the sigmoidal nonlinearity is applied. With one output unit per class, and $d_{ck} = 1$ for training vectors from class c, $d_{ck} = 0$ otherwise, minimizing $E_{\sigma-LMS}$ produces outputs that approximate the Bayesian *a posteriori* probabilities. An input vector is then classified according to the largest output value. However, for practical applications (finite training sets and networks with limited functional complexity), $E_{\sigma-LMS}$ function is not guaranteed to minimize misclassification error (Barnard & Casasent 1989).

A more natural energy function for classification simply counts the number of training vectors that the network misclassifies. The formulation of this counting function varies depending on the output encoding. For a two-class problem, a single output unit suffices, with positive outputs indicating one class and negative outputs indicating the other. A counting function for this network is given by (Szu & Telfer 1991, Telfer & Szu 1992a)

$$E_{MME} = N - \sum_{n=1}^{N} step(d_n o_n), \qquad (2)$$

where d_n is the desired sign of the actual output o_n and $step(z) = 1$ if $z \geq 0$; $step(z) = 0$ otherwise. (Eq. 2 thus uses a sharp membership function; a fuzzy logic version would be an obvious extension.) When the desired sign is the same as that of the actual output o_n, the n-th training vector $\mathbf{x}_n$ is correctly classified, the step function equals 1, and the number of misclassifications E_{MME} is reduced by one. When the desired output sign and actual output sign differ, $\mathbf{x}_n$ is misclassified, the step function equals 0, and E_{MME} is not reduced. To minimize an energy function with gradient descent, the energy function must be differentiable. Although the step function in Eq. 2 is not differentiable, it can be approximated by a sigmoidal function that is gradually steepening. As the magnitudes of the network weights increase, the magnitudes of the network outputs o_n also increase, and the sigmoid behaves more and more like a step function required by Eq. 2.

For multiple classes, if there is one output unit per class and an input is classified based on the largest output, an appropriate counting function, called the Classification Figure of Merit (CFM) (Hampshire & Waibel 1990), is given by

$$E_{CFM} = N - \sum_{n=1}^{N} \sigma(o_{max} - o_{other}), \qquad (3)$$

where o_{max} is the output from the unit that should have the maximum value (corresponding to the training vector's class) and o_{other} is the largest value of the other output units. Here the step function has been replaced by a sigmoid with the above justification. For a correct classification, $o_{max} - o_{other} > 0$ and

$\sigma(o_{max} - o_{other}) \rightarrow 1$, and the number of misclassifications is reduced by 1. For a misclassification, $o_{max} - o_{other} < 0$ and $\sigma(o_{max} - o_{other}) \rightarrow 0$, and the number of misclassifications is not reduced. A proof showing that minimizing E_{CFM} does give the desired result is given in (Hampshire & Pearlmutter 1991).

With multiple classes, the outputs may also be binary encoded labels, in which case the outputs are passed through a threshold rather than a maximum detector. An advantage of binary encoded outputs over one output unit per class is that fewer output units are required. For example, for 16 classes, one output unit per class requires 16 output units, but binary encoded outputs require only 4 output units. In addition, error correcting codes can be used as class labels. For example, a Hamming code (Lin & Costello 1983) with 7 output units can encode 16 classes and correct a single error in the output units. Such an error correcting approach increases classification accuracy and has been shown to improve associative memory performance (Liebowitz & Casasent 1986, Casasent & Telfer 1992). A new MME energy function to minimize misclassification error for binary encoded outputs is given by

$$E_{MME} = N - \sum_{n=1}^{N} \sigma[\sum_{k=1}^{K} \sigma(d_{nk} o_{nk}) - K + 0.5]. \quad (4)$$

The summation over k equals the number of correct output units for the n-th training vector. If all are correct, the summation equals K, and the outer sigmoid becomes 1, which reduces the number of incorrect misclassifications by 1. If there are one or output errors, the summation over k equals at most $K - 1$ (for a single output error) and the outer sigmoid becomes 0, and the misclassification count is not reduced. Note that in this case of multiple classes, E_{MME} must determine from all the output units whether a classification is correct or not. It is not sufficient to simply sum the errors from each output unit individually by summing Eq. 2 over multiple classes.

2-D EXAMPLE

Before considering the Thematic Mapper data, we consider a simpler two-class example of synthetic data with two features. This allows the class boundaries to be easily visualized to provide insight into LMS and MME energy functions. Since the data set is much smaller than the Thematic Mapper data, it also allows more detailed study.

Two classes with equal *a priori* probabilities are drawn from concentric circular uniform distributions with radius $\sqrt{2}/2$ (class 1) and 1 (class 2). The Bayes rate (minimum error) is 0.25, with a circular boundary of radius $\sqrt{2}/2$. The training set consists of 1000 vectors from each class and is shown in Figure 1. (The class boundaries shown in Figure 1 will be discussed shortly.) The test set consists of 5000 vectors from each class. The larger test set is needed to increase the confidence levels of the results.

The following study considered L_1 and L_2 norm versions of the two-class E_{MME}. More details are provided elsewhere (Telfer & Szu 1992b). The method described in the formulation section is the L_1 version. Multilayer perceptrons with two layers of weights and varying numbers of hidden units were tested for σ-LMS, MME L1 and MME L2. The procedure was to randomly initialize the weights to values between ± 1, first train each network for 200 iterations (epochs) using σ-LMS, and then using that result as a starting point, train for 800 iterations using the three energy functions. The motivation for the initial 200 iterations was to move the networks into a reasonable area of weight space which could then be tuned further by each energy function. This was found to produce better results than simply starting with each energy function from random weights. Other random weight magnitudes were also tried to ensure that the best results possible from each energy function were being measured. A conjugate gradient method (Fletcher 1987) was used (restart cycle of 5) with a simple inexact line search in implementing the backpropagation algorithm. For each number of hidden units, ten initial sets of random weights were constructed. In an attempt to discount runs that became stuck in local minima, only the run that gave the minimum training set error for each energy function was included in the results.

Figure 2a plots the performance of each energy function vs. number of hidden units. The MME energy functions produce excellent results with only three hidden units, and as more hidden units are added, they descend to essentially identical training set errors of 0.246 for MME L2 and 0.248 for MME L1 with 8 hidden units. Since it was plain that the MME energy functions were reliably finding minimum error networks, their hidden units were not increased beyond 8. For σ-LMS, the training set error also slowly decreased with increasing numbers of hidden units, but consistently remained higher than the MME training set results and the Bayes rate. With 16 hidden units, σ-LMS still gave 0.259 error, over 1%

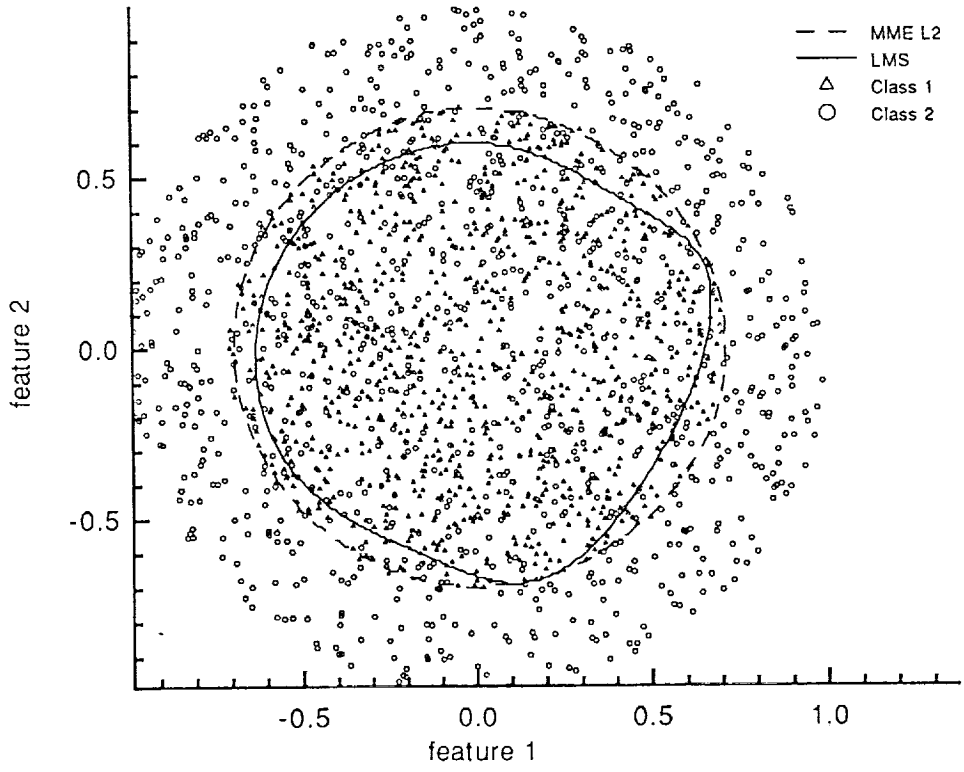


Figure 1: Training set for 2-D case with class boundaries found by σ-LMS and MME L2 networks with four hidden units.

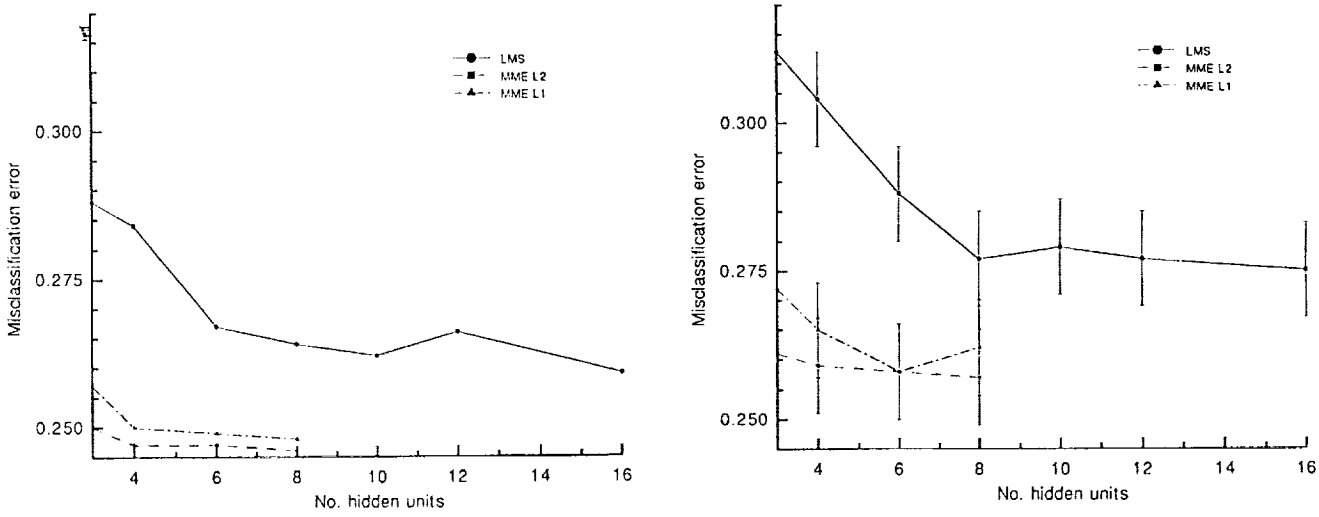


Figure 2: (a) Training set and (b) test set results for different energy functions vs. number of hidden units.

higher than for the MME energy functions. For an infinitely large training set, σ-LMS would converge to the Bayes rate, but this result does not hold for a finite training set.

These results are also reflected in the MME L2 and σ-LMS class boundaries with four hidden units, plotted in Figure 1. The MME L2 boundary is clearly almost exactly the desired circle, while the σ-LMS boundary is consistently inside the optimal boundary.

Of course, the more important question is how the networks performed on the test set. These results are plotted in Figure 2b with 95% confidence intervals (Highleyman 1962). The test set errors are all higher than the respective training set results as expected. The MME test set results are still lower than the σ-LMS results, and the results are statistically significant (although there is a slight overlap between the MME L2 and σ-LMS results at 8 hidden units, even this is still significant with a high but less than 95% confidence level). Even with 16 hidden units, the σ-LMS result is still significantly (in the statistical sense) worse than all but 4 of all 8 MME results with 8 or fewer hidden units. Thus, for this example, σ-LMS requires roughly five times the number of hidden units of the MME energy functions (16 vs. 3) to give equal test set performance.

LANDSAT EXAMPLE

Description of Data

Landsat-4 Thematic Mapper (TM) data taken in July 1982 over an area in the vicinity of Washington, D.C. were used in this study. The TM is a 7-band instrument, with spectral coverages 0.45-0.52 (TM1), 0.52-0.60 (TM2), 0.63-0.69 (TM3), 0.76-0.90 (TM4), 1.55-1.75 (TM5), 10.40-12.50 (TM6), and 2.08-2.35 (TM7). The ground Instantaneous Field-of-View (IFOV) is 30m except for the thermal bands (TM6), which is 120m. As the infrared and the thermal bands had not yet cooled off after launch, only the first four bands are usable.

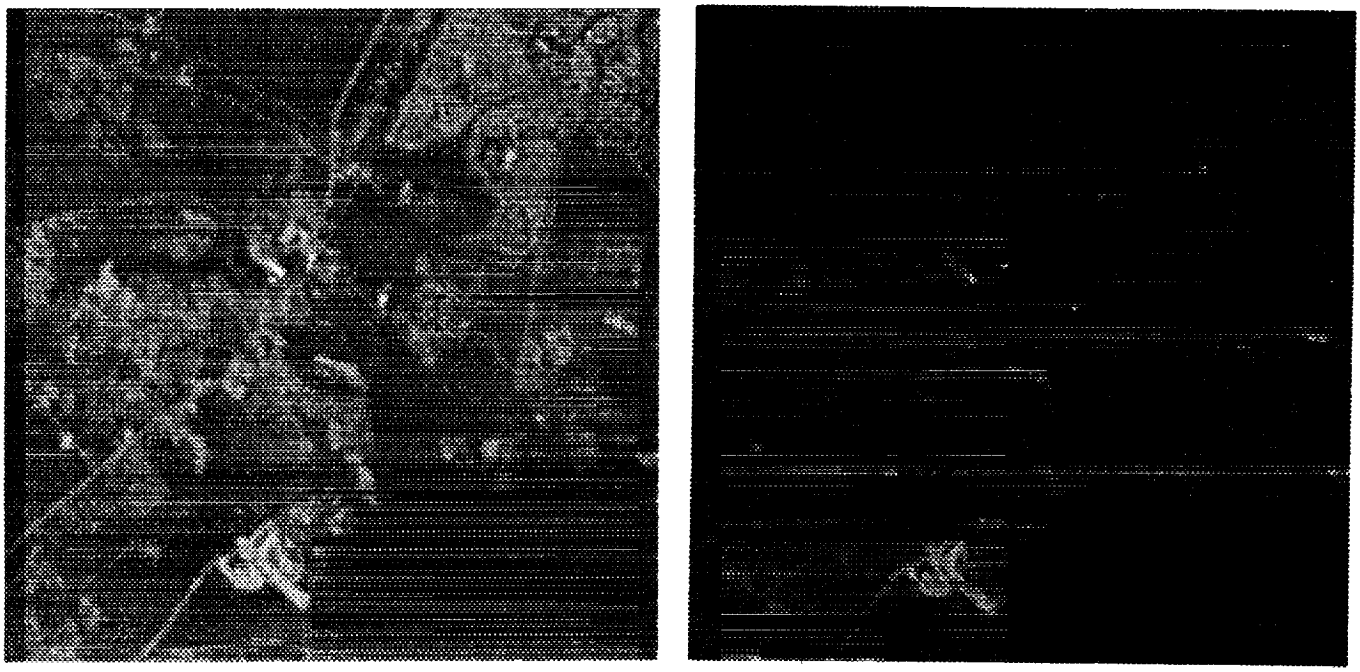
The ground truth consists of 17 categories, and were obtained through photointerpretation of color infrared aerial photographs and subsequent field visits (Williams et al. 1984). Specifically, the categories are (1) water, (2) miscellaneous crops, (3) standing corn, (4) corn stubble, (5) shrubland, (6) grassland or pasture, (7) soybeans, (8) bare soil/cleared land, (9) mostly hardwood dense canopy, (10) mostly hardwood less dense, (11) mostly conifer, (12) mixed wood, (13) asphalt, (14) single-family residential area, (15) multi-family residential area, (16) industrial or commercial area, and (17) bare soil/plowed fields.

In general, ground truth contains information categories instead of spectral categories. As the IFOV is broad enough to cover multiple ground categories, there are natural overlaps among the spectral signatures for these categories. Since the neural networks in this study perform classifications based on spectral data alone, whether the information categories correspond to distinct spectral categories should be examined, in order to estimate the intrinsic discriminability among the categories.

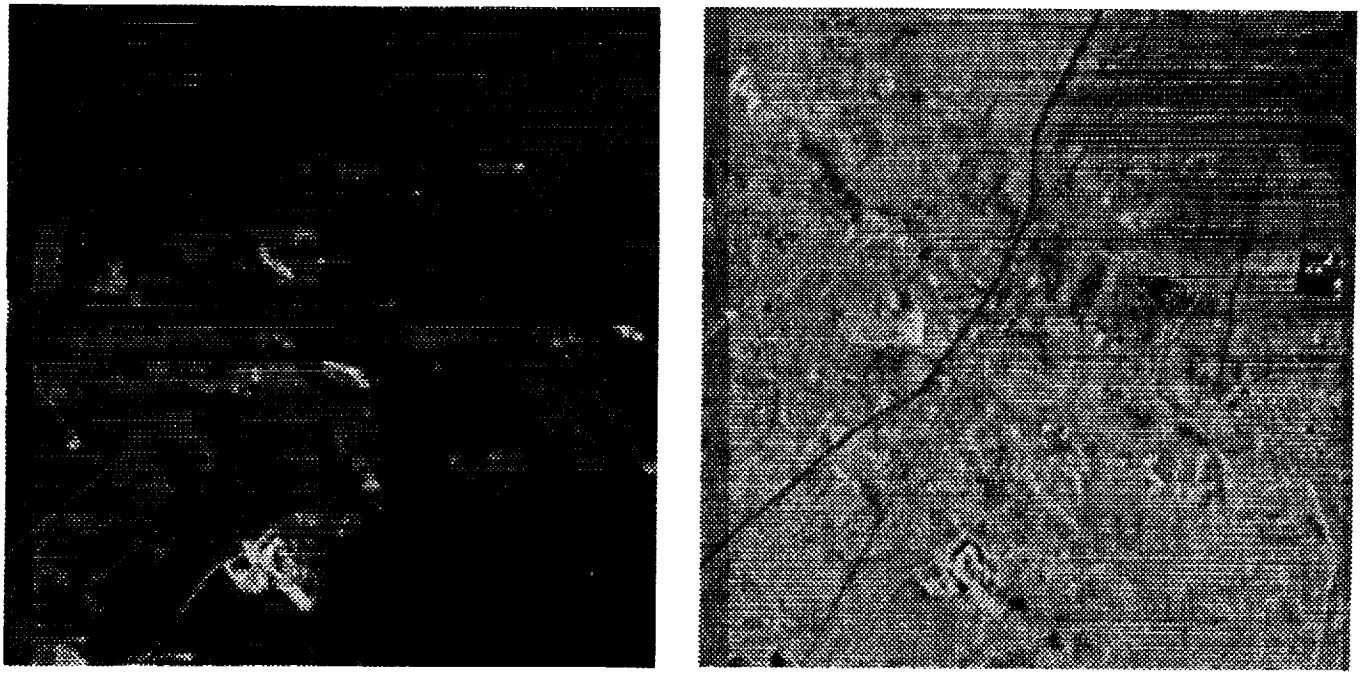
To achieve this objective, the spectral signatures for all categories are computed. The signatures consist of mean vectors and covariance matrices. A number of measures, such as divergence and Mahalanobis distance, could be used to estimate the separability among multi-dimensional clusters. In this study, we compute the ratio of between-class variance to within-class variance along the Fisher optimal discriminant vector (Duda & Hart, 1973). From the ratios, it is concluded that some information categories are heavily overlapped with others, and that the 17 information categories should be combined into 6 categories, following the land use and land cover classification system of Anderson et al. (1976). These six categories are: (1) urban or built-up land, (2) agricultural land, (3) rangeland, (4) forest land, (5) water, and (7) bare soil/cleared land. Notice that there is no Category 6 (wetland) in this data. In Anderson's system, Category 7 is barren land, such as salt flats, beaches, bare rock, etc. Since bare soil/cleared land (Category 17 in the ground truth data) does not exactly fit the definition, the original description in the ground truth is used instead.

To give an idea of the terrain types present, Figure 3 shows the four bands of the 256×256 image (slightly cropped for display purposes). Roads are clearly visible. A housing development is at the upper right. Fields are visible in the center of the image. The dark areas are primarily forest.

The area for which ground truth exists (a roughly 150×150 area in the center of Figure 3) has 21,952 pixels, with pixels placed alternately into training and test sets, giving 10,976 pixels for each. The number of pixels in each class is give in Table 1. Since each pixel contains four spectral bands, each feature vector contains four features, with an additional element set to one to provide a bias term. Each of the four



(a) (b)



(c) (d)

Figure 3: Four bands of Thematic Mapper data: a) TM1, b) TM2, c) TM3, d) TM4.

Class Name	No. Pixels
Urban	2754
Agric	1670
Range	3184
Forest	13781
Water	28
Barc	535

Table 1: Class distribution of Landsat data.

spectral features was normalized to have zero mean and standard deviation of 0.75.

Procedure and Results

Multilayer perceptrons with two layers of weights and twelve hidden units were tested for $E_{\sigma-LMS}$, E_{CFM} and E_{MME}. (Networks with fewer hidden units were also tried but found to perform slightly worse.) For the network structure of one output unit per class, six output units were used, while for binary encoding, three output units were used. (Two of the possible eight codes were unused.) The procedure was to randomly initialize the weights to values between ± 1, first train each network for 500 iterations (epochs) using σ-LMS, and then using that result as a starting point, train for 1000 iterations using the three energy functions. A conjugate gradient method was used (restart cycle of 5) with a simple inexact line search.

The resulting classification accuracies are given in Table 2. We first consider the results for one output per class. Although CFM improved the σ-LMS training set accuracy by 1%, the test set results are identical. The small training set improvement indicates that σ-LMS is finding class boundaries very close to the minimum error boundaries. The excellent σ-LMS performance can be explained by the large training set size and apparently relatively small functional complexity needed to represent the *a posteriori* probabilities in this case.

For the binary coded outputs, the σ-LMS outputs esimate the probabilities that the outputs are 1 given the input. This can be seen to perform worse than MME, which improves accuracy by 2.2% for the training set and 1.2% for the test set. The difference in the test set result is significant with an 88% confidence level. (The 95% confidence level is $\pm 0.75\%$.) There is no statistically significant difference between the two test set results for one output per class and MME

Energy Function	Accuracy (%)		Output Encoding
	Train	Test	
σ-LMS	78.1	78.7	1/class
CFM	79.1	78.7	1/class
σ-LMS	76.4	76.9	binary code
MME	78.6	78.1	binary code

Table 2: Classification accuracies for Landsat data.

binary encoding, but these three results do differ significantly from the σ-LMS binary encoding result. Although the saving in weights by binary encoding is not large in this example, for larger numbers of classes, the savings becomes significant. In addition, the binary encoding performance would be improved by using error correcting codes.

For comparison, a classical nearest neighbor classifier (Duda & Hart, 1973) gave 71.7% test set accuracy. Also, seven previous neural network tests (with various network architectures and sizes) on this data set have given test set accuracies between 71.6% and 78.4% (Kiang 1992, Hwang et al. 1993). Our best result of 78.7% is statistically better than all but one of these previous results (78.4%), and was obtained with a much smaller network - 132 weights for our network vs. about 640 weights required for the radial basis function network giving 78.4%. The fact that this previous result is similar to our best results suggests that this could be the best possible accuracy that can be obtained by classifying single pixels. Further accuracy improvements can be obtained by making use of subpixel information and by classifying based on a neighborhood of pixels. We discuss this in the next section.

There have been neural network based studies (e.g. Bischof et al. 1992) in which classification accuracies are higher than ours. However, it must be pointed out that a direct, fair comparison among these studies may not be possible. As known in remote sensing applications, classification accuracies are highly dependent on the ground types involved, the sensors' resolutions, the seasons when the measurements were taken and the environmental conditions. In general, discrimination among various kinds of vegetation covers is rather difficult.

DISCUSSION OF FUTURE WORK

The 78.7% classification accuracy for single pixel classification should be regarded as a starting point to benchmark further improvements that involve both subpixel and superpixel information. In addition to this work, a method for improving the training set is also discussed.

Since the Thematic Mapper pixel footprint is 30m, the spectra from different landuse types can be mixed in a single pixel. In related work (Shimabukuro & Smith 1991), mixture components are estimated using conventional least squares techniques in order to estimate ages of eucalyptus areas. Neural network apporaches remain to be tested. Since a neural network trained by LMS estimates *a posteriori* probabilities, these can be used as mixing proportions to provide subpixel classification results. For example, it was observed in the LMS classification results described above that many of the pixels along a road passing through forest had large outputs corresponding to both urban (manmade) and forest. Rather than classifying the pixel as urban (road) *or* forest based on only the single largest output, it seems more appropriate to classify the pixel as a certain fraction urban/road *and* a certain fraction forest based on the two largest mixing components. Simply classifying based on the largest output was observed to create many discrepancies with ground truth. For example, the groundtruth marks only discontinuous stretches of the road as urban and the rest as forest. The LMS neural network classifies (based on largest output) the entire stretch of road as urban, but also has a high second largest output for forest. Thus, making use of subpixel mixtures should improve results. Mixture information provides general information about a pixel, but does not indicate the physical region within the pixel occupied by a particular ground type. Super-resolution theory appears promising for physically locating ground types within pixels based on the classifications of nearby pixels.

Conversely, since land use occurs in patches larger than the 30m pixel size, it seems clear that information from neighboring pixels should also increase classification accuracy. Several such ideas for making use of context have been tested with conventional classifiers (Mohn et al. 1987 [tests several prior approaches], Lee & Philpot 1991, Jeon & Landgrebe 1992) and a neural network approach (Bischof et al. 1992). The neural network approach combines spectra from the pixel to be classified and from neighboring pixels into a single feature vector. The neural network then learns from the training set how much weight should be placed on information from neighboring pixels in classifying the central pixel. Bischof et. al. demonstrated a 5% improvement with this method vs. single pixel classification. We are currently testing this contextual technique with our new MME energy functions. A two-pass hybrid spectral/spatial approach is also planned to overcome projection registration and distortion problems.

Lastly, editing the training set should also help improve results. As noted elsewhere (Williams et al. 1984), any minor errors registering groundtruth with the Thematic Mapper data could result in mislabeled samples. Therefore, training samples near class boundaries in the image should be deleted.

ACKNOWLEDGEMENT

The support of this work by by the NSWCDDWO Independent Research Program and an Office of Naval Research Young Navy Scientist Award is gratefully acknowledged. The authors thank Gerald Dobeck for fruitful discussions involving the MME energy function for binary encoded outputs.

REFERENCES

Barnard, E., D. Casasent (1989), "A Comparison Between Criterion Functions for Linear Classifiers. *IEEE Trans. Systems Man and Cybernetics*, Vol. 19, pp. 1030–1041.

Barron, A., T. Cover (1991), "Minimum Complexity Density Estimation," *IEEE Trans. Information Theory*, Vol. 37, pp. 1034–1054.

Benediktsson, J.A., P.H. Swain, O.K. Ersoy (1990), "Neural Network Approaches Versus Statistical Methods in Classification of Multisource Remote Sensing Data," *IEEE Trans. Geoscience and Remote Sensing*, Vol. 28, p. 540.

Bischof, H., W. Schneider, A.J. Pinz (1992), "Multispectral Classification of Landsat Images using Neural Networks," *IEEE Trans. Geoscience Remote Sensing*, Vol. 30, pp. 482–490.

Campbell, W.J., S.E. Hill, R.F. Cromp (1989), "Automatic Labeling and Characterization of Objects Using Artificial Neural Networks," *Telematics and Information*, Vol. 6, p. 259.

Casasent, D., B. Telfer (1992), "High Capacity Pattern Recognition Associative Processors," *Neural Networks*, Vol. 5, pp. 687–698.

Decatur, S.E. (1989), "Application of Neural Networks to Terrain Classification," *Proc. International Joint Conf. Neural Networks*, p. I-284.

Duda, R., P. Hart, (1973), *Pattern Classification and Scene Analysis*, New York: John Wiley and Sons.

Fletcher, R. (1987), *Practical Methods of Optimization*, New York: John Wiley and Sons.

Hampshire, J., A. Waibel (1990), A Novel Objective Function for Improved Phoneme Recognition Using Time-Delay Neural Networks. *IEEE Trans. Neural Networks*, Vol. 1, pp. 216–228.

Hampshire, J., B. Pearlmutter (1991), Equivalence Proofs for Multi-layer Perceptron Classifiers and the Bayesian Discriminant Function. In Touretzky, Elman, Sejnowski, and Hinton, eds., *Proceedings of the 1990 Connectionist Models Summer School*, San Mateo, CA: Morgan-Kaufmann, pp. 159-172.

Highleyman, W. (1962), "The Design and Analysis of Pattern Recognition Experiments, *Bell System Technical Journal*, Vol. 41, pp. 723-744.

Hwang, J.N., S.R. Lay, R.K. Kiang (1993), "Robust Construction Neural Networks for Classification of Remotely Sensed Data," submitted to *World Conference on Neural Networks*, Portland OR.

Jeon, B., D.A. Landgrebe, "Classification with Spatio-Temporal Interpixel Class Dependency Contexts," *IEEE Trans. Geoscience Remote Sensing*, Vol. 30, 663–672.

Kiang, R.K. (1992), "Classification of Remotely Sensed Data using OCR-Inspired Neural Network Techniques," *Proc. 1992 International Geoscience and Remote Sensing Symposium*, Houston TX, pp. 1081–1083.

Lee, J.-H., W.D. Philpot (1991), "Spectral Texture Pattern Matching: A Classifier for Digital Imagery," Vol. 29, pp. 545–554.

Liebowitz, S., D. Casasent (1986), "Error Correction Coding in an Associative Processor," *Applied Optics*, Vol. 26, pp. 999-1006.

S. Lin & D.J. Costello (1983), *Error Control Coding: Fundamentals and Applications*, Englewood Cliffs, NJ: Prentice-Hall.

Liu, Z.K., J.Y. Xiao (1991), "Classification of Remotely Sensed Image Data Using Artificial Neural Networks," *Int. J. Remote Sensing*, Vol. 12, pp. 2433-2438.

Mohn, E., N.J. Hjort, G.O. Storvik, "A Simulation Study of Some Contextual Classification Methods for Remotely Sensed Data," *IEEE Trans. Geoscience Remote Sensing*, Vol. 25, pp. 796–804.

Richard, M., R. Lippmann (1991), "Neural Network Classifiers Estimate Bayesian a posteriori Probabilities," *Neural Computation*, Vol. 3, pp. 461–483.

Rumelhart, D.E., G.E. Hinton, R.J. Williams (1986), "Learning Representations by Back-Propagating Errors," *Nature*, Vo. 323, pp. 533–536.

Shimabukuro, Y.E., J.A. Smith (1991), "The Least-Squares Mixing Models to Generate Fraction Images Derived from Remote Sensing Multispectral Data," *IEEE Trans. Geoscience Remote Sensing*, Vol. 29, pp. 16–20.

Szu, H., B. Telfer (1991), "Minimum Misclassification Error Performance Measure for Layered Networks of Artificial Fuzzy Neurons," *Proc. International Joint Conference on Neural Networks*, Vol. II, p. A-916.

Telfer, B. H. Szu (1992a), "Implementing the Minimum-Misclassification-Error Energy Function for Target Recognition," *Proc. International Joint Conference on Neural Networks-Baltimore*, vol. IV, pp. 214–219.

Telfer, B., H. Szu (1992b), "Energy Functions for Minimizing Misclassification Error with Minimum-Complexity Networks," submitted to *Neural Networks*.

Werbos, P. (1974), "Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences," Ph.D. thesis in applied mathematics, Harvard University.

Williams, D.L. et. al. (1984), "A Statistical Evaluation of the Advantages of Landsat Thematic Mapper Data in Comparison to MSS Data," *IEEE Trans. Geoscience and Remote Sensing*, Vol. 22, pp. 294–302.

Knowledge Engineering