The Second International Conference "Problems of Cybernetics and Informatics"
September 10-12, 2008, Baku, Azerbaijan. Section #2 "Intellectual Systems"
www.pci2008.science.az/2/04.pdf

INVESTIGATION OF THE DESIGN AND USE OF ARTIFICIAL NEURAL NETWORKS IN THE CLASSIFICATION OF REMOTELY SENSED IMAGES

Hikmat Hasanov[1], Tofig Suleymanov[2], Elena Gambarova[3]

[1)] Institute for Aerospace Informatics, National Aerospace Agency, Baku, Azerbaijan
[2)] Institute for Ecology, National Aerospace Agency, Baku, Azerbaijan
[3)] Institute for Aerospace Informatics, National Aerospace Agency, Baku, Azerbaijan,
YLebedik@azuni.net

Abstract

Artificial neural networks (ANNs) are used for rare vegetation communities' classification using remotely sensed data. Training of a neural network requires that the user specifies the network structure and sets the learning parameters. In this study, the optimum design of ANN's for classification of remotely sensed images data is investigated. Heuristics proposed by a number of researchers to determine the optimum values of network parameters are compared using datasets. We use test datasets in a series of experiments that evaluate the effects on network performance (measured in terms of the Mean Square Error, Correlation coefficient and Error per element of the neural network values) of different choices of network size and structure, network parameters, training samples size.

Keywords*: Neural networks, remote sensing, classification, IKONOS imagery*

1. Data used and methodology

Two IKONOS images acquired in July 2005 and June 2006 were used for the delineation of 12 rare vegetation communities and soil types. The study area was about a 110 km^2 region of flat land located in the south-east of the Azerbaijan Republic. The images being used were pan-sharpened multi-spectral images with a spatial resolution of 1m. It was defined test sites for 12 classes on the both images where training and test samples were gathered from [1]. Artificial neural networks (ANNs) are used in the classification of remotely sensed data.

And backpropagation learning algorithm, also called the generalized delta rule, was an iterative gradient descent training procedure.

A specialized GIS was used as software environment for performing of workflow comprising of jobs connected with collecting of samples, hosting of classifier training and production software as well as classification results analysis.

2. Training of Multilayer Perceptron (MLP) classifier

The MLP classifiers training process depends on a set of parameters. The value of these parameters can be varied to different extent (degree) depending on different conditions of experiment. There are the following parameters which were varied during the experiment:

- Number of training samples;
- Stopping criterion for the training process (number of iterations);
- Number of input nodes;
- Learning rate and momentum;
- Number of hidden layers;
- Number of hidden nodes;
- Type of an activation function

Figure 1 depicts the topological structure of the back-propagation neural network in which neurons are arranged into three layers, (i.e., input layer, hidden layer and output layer).

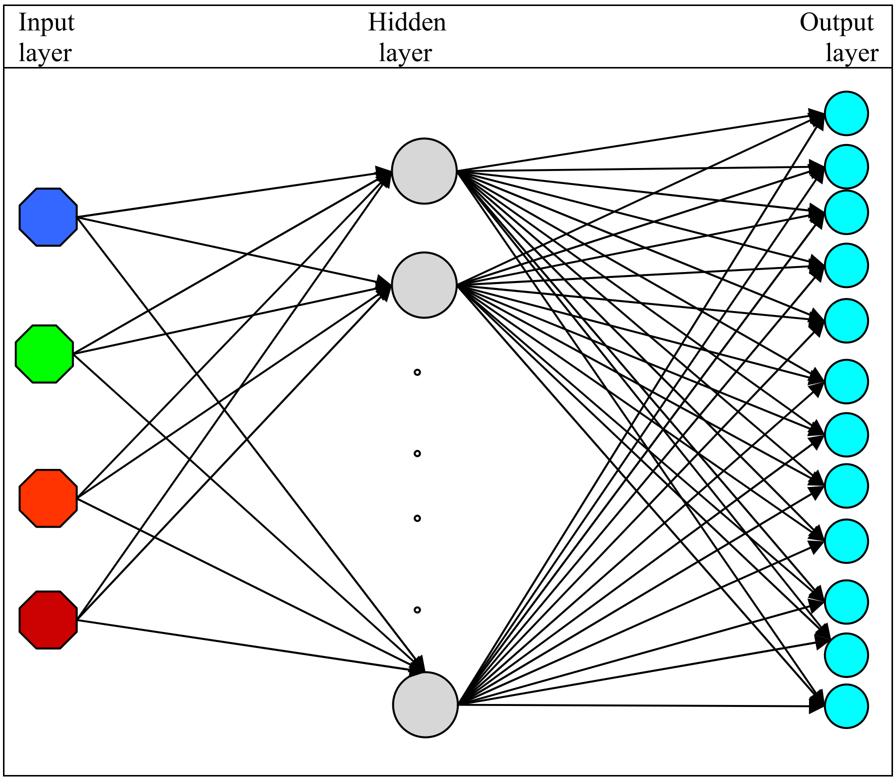

Figure 1. The topological structure of the backpropagation neural network used in the study.

The input layer consisted of 4 neurons, corresponding to four spectral channels of IKONOS satellite scanner: we used the red, green, blue, and near infrared (NIR) channel. The hidden layer had 25 neurons and the output layer had 12 neurons.

An activation function was hyperbolic tangent.
The backpropagation algorithm was used for neural network training.
A network structure of 4-25-12 was trained with the parameters listed in Table 1.

Table 1. Optimum setting of network structure and learning parameters

Parameters	**Choice**
Initial weight range	[0, 0.05]
Number of input nodes	4
Number of hidden layers	1
Number of hidden nodes	25
Learning rate between input and hidden layers	0.5
Momentum term between input and hidden layers	0.7
Learning rate between hidden and output layers	0.25
Momentum term between hidden and output layers	0.7
Type of activation function	hyperbolic tangent
Error threshold er_{thresh}	0.01

The number of training samples employed at the learning stage has a significant impact on the performance of any classifier. In most remote sensing studies the size of training samples is limited, but the design of a neural network is partly based on the number of training samples.

There have been several attempts to estimate the optimum number N_T of training samples [2], [3], and [4]:

$$N_T = 5 \times N_W \tag{1}$$

$$N_T = 10 \times N_W \tag{2}$$

$$N_T = 30 \times N_I \times (N_I + 1) \tag{3}$$

Where:

N_I is the number of input nodes. In our case, N_I is the number of spectral channels;

N_O is the number of output nodes or the number of classes;

N_H is the number of hidden nodes.

Several strategies and heuristics (pruning, constructive methods and hybrid techniques coupling both methods) have been suggested to estimate the optimum number of hidden layer nodes by the following formulas [5], [6], and [7]:

$$N_H = 2 \times N_I \quad \text{or} \quad N_H = 3 \times N_I \tag{5}$$

$$N_H = 2 \times N_I + 1 \tag{6}$$

$$N_H = (N_I + N_O)/2 \tag{7}$$

N_W is the total number of weights in the network which was suggested by [2] as:

$$N_W = N_H \times (N_I + N_O) \tag{9}$$

For determination the optimal number N_T of training samples we used the heuristic suggesting the use of training samples for each weight in the network. The results show that the heuristics proposed by [2] and [3] are good choices.

In our case, for N_I = 4 (the number of spectral channels), N_O = 12 (the number of classes), N_H = 25 (the number of hidden nodes), the optimal number N_T of training samples was determined for selected intervals [3420, 24000].

The average number of samples (for training and test sets) is about 10000.

For evaluation of training process we used the following quality parameters:

- The Mean Square Error (*MSE)*;

Having reached *threshold* set for the *MSE* level of 0.01 the training process was stopped;

- The Correlation coefficient (*r*), which reflects the degree of correlation between directions of changing of real and desired outputs of the neural network;
- $\%Error$ - Error per element of the neural network

During the training process, all learning parameters were kept constant. Training networks were saved and their performances evaluated using a test dataset.

When we reached the maximal number of iterations and the training process was stopped¸ we compiled a confusion matrix with the results of recognition of samples from training sets (so called self-testing procedure).

We estimated the common degree of correctness (CDC) by the following formula:

$$CDC = 100\% \times (N_{CCS} / N_{total}) \qquad (10)$$

Where,
N_{CCS} is the number of correctly classified samples;
N_{total} is the total number of samples

The confusion matrix computed on results of the classifier training showed that:

CDC = 96.37% (for training samples)
CDC = 91.45% (for test samples)

Summarizing the classifier testing results, we conclude that beside some realized uncertainties in recognizing of test samples and non-ideal values of quality parameters being received, in the whole, the neural classifier has been trained properly. In case of arising difficulties in production period, additional geo-spatial data (DTM and its derivable, other topographical, hydrological data as well as land use information and etc.) could be involved into the process and the problems of recognition of objects would be solved.

Literature

1. E. M. Gambarova. The practical principles of neural classifiers training for object recognition on multi-spectral high resolution satellite images. *Bulletin of Azerbaijan National Aerospace Agency*, Tom 10, №3-4, (2007), p. 23-34.
2. T. Kavzoglu and P.M. Mather. The use of backpropagating artificial neural networks in land cover classification, *INT. J. Remote Sensing*, 24, (2003), p. 4907–4938.
3. C.C. Klimasauskas. Applying neural networks, *In Neural Networks in Finance and Investing*, edited by R. R. Trippi and E. Turban (Cambridge: Probus), (1993), p. 47–72.
4. D. R. Hush. Classification with neural networks: a performance analysis, Proceedings of the *IEEE International Conference on Systems Engineering*, Dayton, Ohio, USA, (1989), p. 277–280.
5. I. Kanellopoulos and G.G. Wilkinson. Strategies and best practice for neural network image classification, *International Journal of Remote Sensing*, 18, (1997), p.711–725.
6. R.R. Hecht-Nielsen. Kolmogorov's mapping neural network existence theorem, *Proceedings of the First IEEE International Conference on Neural Networks*, San Diego, (1987), p.11–14.
7. B. Ripley. Statistical aspects of neural networks, *In Networks and Chaos – Statistical and Probabilistic Aspects,* (London: Chapman & Hall), (1993), p. 40–123.