

USE OF NEURAL NETWORKS FOR LAND COVER CLASSIFICATION FROM REMOTELY SENSED IMAGERY

L. Zhou *, X. Yang

Department of Geography, Florida State University, 113 Collegiate Loop, Tallahassee, FL 32306, USA
- (lz06c, xyang)@fsu.edu

Commission VII, WG VII/4

KEY WORDS: Land Cover Classification, Neural Networks, Topology, Activation Function, Neural Training, Accuracy,

ABSTRACT:

Remote sensing has been increasingly used to derive land cover information by manual interpretation or automated classification. As promising automated classifiers, artificial neural networks are difficult to parameterize, thus challenging their applicability. Previous studies investigated the impacts of some external factors, such as input data and training samples, upon neural network classification, but paid relatively little attention on how to appropriately parameterize their internal parameters. In this paper, we report the result of a pilot project aiming to develop some guides for parameterizing the multi-layer-perceptron (MLP) feed-forward back-propagation neural networks. The internal parameters we consider include the number of hidden layers, activation function, learning rate, momentum, threshold, and number of iterations. We choose part of the Atlanta metropolitan area, Georgia in the U.S.A as the test site where the landscape mosaic is characterized by several decades of rapid urban growth. We carefully configure 59 neural networks models with different internal parameters settings. We train these models with the same sample dataset and use each of them to map land cover information from a Landsat Enhanced Thematic Mapper Plus (ETM+) image. We compute the overall classification accuracy for each derived map by using the identical reference data. We found that the four internal parameters, namely, activation function, learning rate, momentum, and number of iterations, significantly affect classification accuracy; while the other two parameters, namely, threshold and number of hidden layers, also impact the classification performance. Finally, we propose a guideline that can help parameterize the MLP neural networks, and thus our study should help promote the operational use of neural networks for land cover classification.

1. INTRODUCTION

Land cover refers to the physical and biological cover at the surface of the earth, including water, vegetation, bare ground, man-made structures, etc. In recent years, land cover is found to be closely associated with our environments and ecosystems, such as global and local climate, hydrologic circle, pollutions, biodiversity, and soil erosions. Due to its technological robustness and cost-effectiveness, remote sensing has been increasingly used to derive land cover information through either manual interpretation or automated classification. The latter is more desirable but can be less effective for classifying heterogeneous landscapes. Artificial neural networks are commonly conceived to have the capability of improving automated classification accuracy due to their distributed structure and strong capability of handling complex phenomena.

Artificial neural networks are the models based on biological neural networks. They consist of an interconnected group of neurons. Each neuron contains a single computation process. The multi-layer-perceptron (MLP) feed-forward back-propagation neural networks are most popular in practice due to their easiness to understand and implement. A typical structure of a MLP neural network is comprised of neurons and their links in a layered structure. The input to a neuron in such a network is the weighted sum of the outputs of neurons at the previous layer. Each neuron at both hidden and output layers contains a single process which is to transform the input using a linear or non-linear function, namely, activation function.

Many experiments have shown that MLP neural networks are more accurate for land cover classification than traditional statistical methods (Bischof *et al.*, 1992; Civco, 1993; Serpico *et al.*, 1996; Chen *et al.*, 1997). However, their applicability has been challenged by the complexity of neural networks parameterization. A number of factors need to be considered when parameterizing neural networks, which include input data, training samples, output settings, networks architecture, initial weights, and training parameters. Most of the previous studies targeted several external factors, such as input data and training samples. Several studies investigated some internal parameters, such as the number of hidden layers (Civco, 1993; Kanellopoulos and Wilkinson, 1997; Foody and Arora, 1997; Kavzoglu and Mather, 2003), activation function (Shupe and Marsh, 2004), and the combination of training rate and momentum (Kavzoglu and Mather, 2003). However, their findings are not consistent. Further research is needed to comprehensively investigate the possible impacts of these internal parameters upon the performance of land cover classification.

Through the analysis of MLP neural networks topology and back-propagation training method, the internal parameters that could potentially affect classification accuracy include number of hidden layer and neurons, activation function, learning rate, momentum, threshold, and number of iterations. The effect of the number of neurons has been investigated elsewhere (*e.g.* Foody and Arora, 1997; Kanellopoulos and Wilinon, 1997;

* Corresponding author

Shupe and Marsh, 2004), and our pilot study concentrated on the other six parameters.

The objective of our study was to develop a guideline for setting the internal parameters of the MLP neural networks trained by the back-propagation method. Firstly, we chose an appropriate study area and identified a land cover classification scheme. Secondly, we carefully configured 59 neural networks models with different internal parameters settings and trained each of them using the same sample dataset. We used each of these models to classify an ETM+ image covering the study area, and assess the overall classification accuracy for each land cover map by using the identical reference dataset. Finally, we analyzed the impact of each parameter and proposed a guideline for parameterizing the MLP neural networks trained by the back-propagation method.

2. RESEARCH METHODOLOGY

Optimum internal parameters setting can be obtained by separately investigating the impact of each individual parameter on classification accuracy. We carefully configured a total of 59 neural networks models by altering the value of one parameter while holding other parameters unchanged. These models have been used to derive land cover information from remotely sensed data. The effect of each parameter can be evaluated in term of classification accuracy. Figure 1 illustrates the working procedural route.

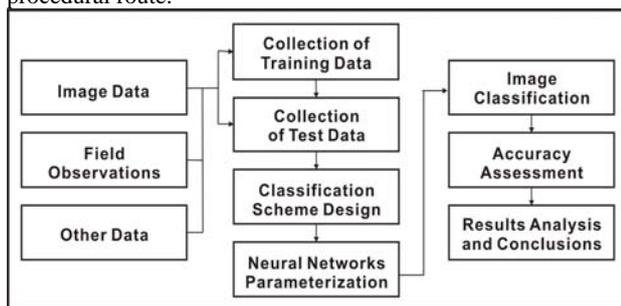


Figure 1. Flowchart of the working procedural route.

2.1 Study Area and Data

Our test site is located in the north-eastern Atlanta metropolitan area, Georgia in the U.S.A (Figure 2). This study area, mainly covering Gwinnet County, Georgia, has been experiencing rapid urban expansion in recent decades (Yang, 2002). The existence of a large amount of transitional land uses (such as the residential areas under construction) challenges the applicability of automated classification. It is hoped that appropriately parameterized neural networks can help improve image classification in such a complex environment.

An ETM+ image obtained on 9 September 1999 was used to derive land cover information. This image had already been accurately rectified and geo-referenced to a UTM map projection (Zone 16), a NAD 83 horizontal datum, and the GRS 1980 ellipsoid. Therefore, geometric rectification was not attempted. To conduct land cover classification by neural networks, seven spectral data channels (except the thermal band) were included as the inputs to each of the neural networks model. The spatial resolution ranges from 15 m for the panchromatic band to 28.5 m for the other five bands. We did not use the thermal band since it has much lower spatial resolution (60*60m) comparing with other bands. The nearest-

neighbor resampling method was adopted to avoid changing the original pixel values of the image data. The original image was resampled to 15 meters.

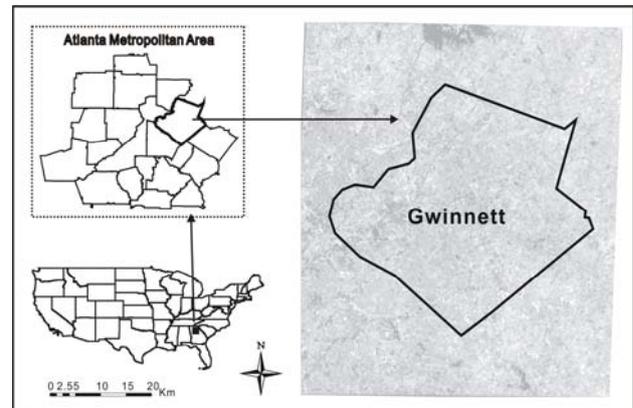


Figure 2. Location of the study area and the image coverage.

In addition to the ETM image, we also collected two types of reference data through field observations and the use of high-resolution images from the Google Earth. The Google Earth provides remote sensor images with spatial resolution as high as 2.8m or better for the test site. With the use of the high-resolution images from the Google Earth, we developed a sample dataset that was randomly spitted into the training subset and the test subset. Field observations were used to supplement or confirm the interpretation from the Google Earth images, particularly for the test samples.

2.2 Land Cover Classification Scheme

The land-cover classification scheme originally designed by Yang (2002) was adopted here. This system includes six major classes: high-density urban use, low-density urban use, exposed land, cropland/grassland, forest, and water.

No.	Class	Description
1	High-density urban use	Approximately 80 to 100 percent construction materials; typically commercial and industrial buildings with large open roofs, large open transportation facilities, and high-density residential areas in the city cores
2	Low-density urban use	Approximately 50 to 80 percent construction materials; often residential development including mostly single/multiple family houses and public rental housing estates as well as local roads and small open spaces in a residential area; with certain amount of vegetation cover (up to 20 percent)
3	Exposed land	Areas of sparse vegetation cover (less than 20 percent), including clear-cuts, quarries, barren rock or sand along river/stream beaches
4	Cropland/grassland	Golf courses, lawns, city parks, crop fields, and pasture
5	Forest	Deciduous, coniferous, and mixed forests and orchards
6	Water	Streams, rivers, lakes, and reservoirs

Table 1. Land cover classification scheme (Yang, 2002)

2.3 Parameterizing Neural Networks Models

The general rule here was to alter the value of one parameter at one time while holding the other unchanged. In this way, the effect of each parameter can be evaluated separately. For example, to investigate the impact of number of hidden layers, five neural networks models were configured with the number of hidden layers ranging from 0 to 4 while holding the other parameters unchanged.

Table 2 shows the model configurations. Note that models 1 to 43 were used to assess the impacts of the internal parameters excluding activation function. In addition, we considered the two widely used non-linear activation functions, namely, log-sigmoid function and tan-sigmoid function; their impacts can be assessed by comparing models 44 to 53 with models 13 to 22. Models 54 to 59 were used to assess the impact of number of iterations upon the performance of the tan-sigmoid function.

2.4 Image Classification and Accuracy Assessment

The training dataset contains 250 samples for each land cover class; the test dataset has 50 samples for each class that were selected by using the stratified random scheme. Note that the identification of each dataset involved the use of the high-resolution images from the Google Earth as the reference.

The 59 neural networks models were trained using the identical training dataset. The training process for most of these models converged except the fifth model with four hidden layers. The training root-mean-square (RMS) errors for these models largely ranged from 0.26 to 0.40; four of them were larger than 0.40. No obvious relationship between the classification accuracy and the training RMS errors had been observed.

A total of 59 land cover classification maps were produced from the ETM+ image by using the trained models. With the test dataset as the reference, the accuracies of these maps were assessed using the thematic accuracy assessment method (see Jensen 2005). An error matrix was generated for each map and the overall classification accuracy was computed (Table 2).

3. RESULTS AND DISCUSSION

Based on Table 2 and Figure 3, it is clear that the accuracies of land cover classification vary across different neural models associated with different internal parameter settings. In particular, the four internal parameters, namely, activation function, training rate, momentum, and number of iterations, are found to highly affect the classification accuracy.

Based on Figure 3a, it is clear that increasing the number of hidden layer does not help improve classification accuracy; the neural network model with one hidden layer shows the best overall accuracy. This observation is consistent with the results obtained by Kanellopoulous and Wilkinson (1997). When the number of hidden layers increased to four, training of the model failed to converge. Based on this experiment, we suggest that neural networks with one hidden layer are suitable for multi-spectral image classification with a few output classes. However, more complex topology may need when a large number of input data are used or when many different classes are to be generated.

The models configured with the tan-sigmoid function had much lower classification accuracies than those with the log-sigmoid function (Figure 3b). The models with the tan-sigmoid function did not vary greatly in their classification accuracies as the number of iterations increased (Table 2).

No.	Internal Parameters*						OA(%)
	HL	AF	LR	MO	TH	IT	
1	0	Log	0.01	0.6	0.2	1000	79.33
2	1	Log	0.01	0.6	0.2	1000	80.00
3	2	Log	0.01	0.6	0.2	1000	77.67
4	3	Log	0.01	0.6	0.2	1000	66.67
5	4	Log	0.01	0.6	0.2	1000	N/A
6	1	Log	0.3	0.6	0.2	1000	69.33
7	1	Log	0.25	0.6	0.2	1000	65.67
8	1	Log	0.2	0.6	0.2	1000	72.67
9	1	Log	0.15	0.6	0.2	1000	71.67
10	1	Log	0.10	0.6	0.2	1000	76.00
11	1	Log	0.05	0.6	0.2	1000	79.33
12	1	Log	0.01	0.6	0.2	1000	80.00
13	1	Log	0.01	0	0.8	1000	72.33
14	1	Log	0.01	0.1	0.8	1000	73.00
15	1	Log	0.01	0.2	0.8	1000	75.67
16	1	Log	0.01	0.3	0.8	1000	75.67
17	1	Log	0.01	0.4	0.8	1000	76.00
18	1	Log	0.01	0.5	0.8	1000	76.67
19	1	Log	0.01	0.6	0.8	1000	79.00
20	1	Log	0.01	0.7	0.8	1000	78.00
21	1	Log	0.01	0.8	0.8	1000	78.67
22	1	Log	0.01	0.9	0.8	1000	78.33
23	1	Log	0.01	0.6	0	1000	82.00
24	1	Log	0.01	0.6	0.1	1000	79.67
25	1	Log	0.01	0.6	0.2	1000	80.00
26	1	Log	0.01	0.6	0.3	1000	80.33
27	1	Log	0.01	0.6	0.4	1000	80.00
28	1	Log	0.01	0.6	0.5	1000	79.67
29	1	Log	0.01	0.6	0.6	1000	79.00
30	1	Log	0.01	0.6	0.7	1000	78.33
31	1	Log	0.01	0.6	0.8	1000	79.00
32	1	Log	0.01	0.6	0.9	1000	72.00
33	1	Log	0.01	0.6	0.2	400	75.33
34	1	Log	0.01	0.6	0.2	700	78.00
35	1	Log	0.01	0.6	0.2	1000	80.00
36	1	Log	0.01	0.6	0.2	1300	80.67
37	1	Log	0.01	0.6	0.2	1600	81.33
38	1	Log	0.01	0.6	0.2	1900	82.00
39	1	Log	0.01	0.6	0.2	2200	82.67
40	1	Log	0.01	0.6	0.2	2500	83.00
41	1	Log	0.01	0.6	0.2	2800	83.00
42	1	Log	0.01	0.6	0.2	3100	84.00
43	1	Log	0.01	0.6	0.2	3400	83.33
44	1	Tan	0.01	0	0.8	1000	40.33
45	1	Tan	0.01	0.1	0.8	1000	40.33
46	1	Tan	0.01	0.2	0.8	1000	40.67
47	1	Tan	0.01	0.3	0.8	1000	43.00
48	1	Tan	0.01	0.4	0.8	1000	40.67
49	1	Tan	0.01	0.5	0.8	1000	34.67
50	1	Tan	0.01	0.6	0.8	1000	40.33
51	1	Tan	0.01	0.7	0.8	1000	39.00
52	1	Tan	0.01	0.8	0.8	1000	17.33
53	1	Tan	0.01	0.9	0.8	1000	17.33
54	1	Tan	0.01	0.6	0.8	2000	30.00
55	1	Tan	0.01	0.6	0.8	3000	29.33
56	1	Tan	0.01	0.6	0.8	4000	39.67
57	1	Tan	0.01	0.6	0.8	5000	40.33
58	1	Tan	0.01	0.6	0.8	6000	40.67
59	1	Tan	0.01	0.6	0.8	7000	39.67

- HL=Number of Hidden Layers; AF=Activation Function; LR=Learning Rate; MO=Momentum; TH=Threshold; IT=Number of Iterations; OA=Overall Accuracy; Log=Log-sigmoid Function; Tan=Tan-sigmoid Function.

Table 2. List of the neural networks models as related to different internal parameters.

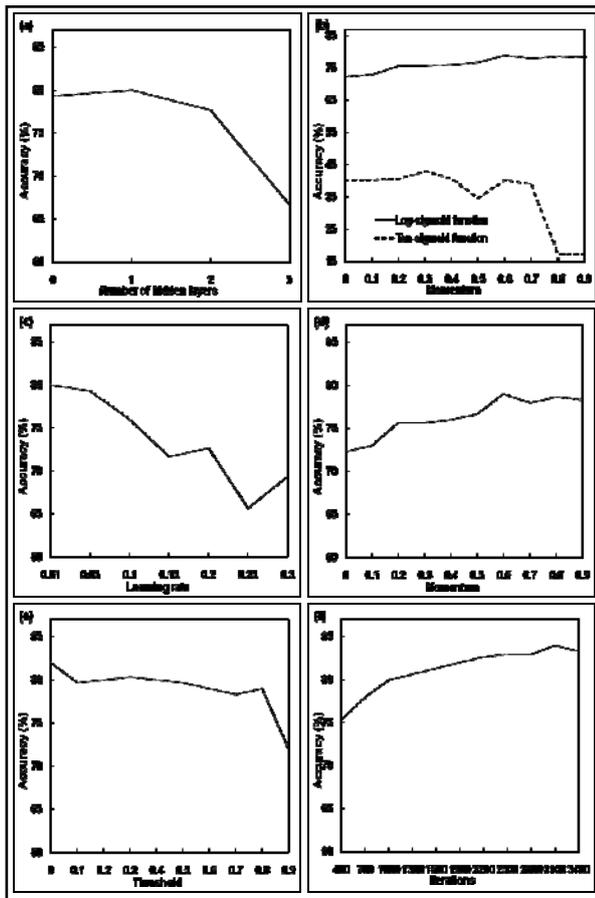


Figure 3. Trends of the overall classification accuracies (Y-axis) as related to different internal parameters (X-axis): (a) number of hidden layers, (b) activation function, (c) learning rate, (d) momentum, (e) threshold, and (f) number of iterations.

By adjusting the learning rate, the classification accuracy increased from 65.7% to 80% (Figure 3c). The classification accuracy increased by more than six percent with the increase of the momentum value (Figure 3d). The classification accuracy increased when the number of iterations increased but this trend slowed down when the number was quite big (see Figure 3f). This indicates the occurrence of over-training which can reduce the generalization capability of neural networks. Apparently, the three training parameters, namely, learning rate, momentum value, and number of iteration, have a significant effect on the classification accuracy. The value of threshold does not seem to be quite influential to the classification accuracy as a large portion of the curve in Figure 3e is quite flat. However, the accuracy plunged when the value of threshold was increased to 0.9 (Figure 3e).

4. SUMMARY AND CONCLUSIONS

In this study, we assessed the impacts of several internal parameters of the MLP neural networks trained by the back-propagation method upon the performance of land cover classification from remote sensor data. A total of 59 models were carefully configured with different parameter settings, which were applied to classify an ETM+ image covering a complex urban/suburban area. The classification accuracies were further compared.

We found that a large number of hidden layers did not help boost classification accuracy. The log-sigmoid function is found to outperform the tan-sigmoid function. We also found that the MLP neural network models have been quite sensitive to the training parameter settings; to optimize the classification performance, we recommend using a small learning rate, moderate momentum, moderate number of iterations, and zero threshold. Therefore, our study should help promote the operational use of neural networks for land cover classification.

REFERENCES

Bischof, H., Schneider, W., Pinz, A.J.,1992. Multispectral classification of Landsat-images using neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 30(3), pp] 482-490.

Chen, K., Yen, S., Tsay, D.,1997. Neural classification of SPOT imagery through integration of intensity and fractal information. *International Journal of Remote Sensing*, 18(4), pp. 763-783.

Civco, D.L.,1993. Artificial neural networks for land-cover classification and mapping. *International Journal of Geographical Information Systems*, 7(2), pp. 173-186.

Foody G.M., Arora, M.K.,1997. An evaluation of some factors affecting the accuracy of classification by an artificial network. *International Journal of Remote Sensing*, 18(4), pp. 799-810.

Jensen, J. R.,2005. *An Introductory Digital Image Processing: A Remote Sensing Perspective* (3rd). Prentice Hall, New Jersey, pp. 296-301.

Kanellopoulos, I., Wilkinson, G.G.,1997. Strategies and best practice for neural network image classification. *International Journal of Remote Sensing*, 18(4), pp. 711-725.

Kavzoglu, T., Mather, P.M.,2003. The use of backpropagating artificial neural networks in land cover classification. *International Journal of Remote Sensing*, 24(23), pp. 4907-4938.

Serpico, S.B., Bruzzone, L., Roli, F.,1996. An experimental comparison of neural and statistical non-parametric algorithms for supervised classification of remote sensing images. *Pattern Recognition Letters*, 17(13), pp. 1331-1341.

Shupe, S.M., Marsh, S.E.,2004. Cover- and density-based vegetation classifications of the Sonoran desert using Landsat TM and ERS-1 SAR imagery. *Remote Sensing of Environment*, 93(2004), pp. 131-149.

Yang, X.,2002. Satellite monitoring of urban spatial growth in the Atlanta metropolitan area. *Photogrammetric Engineering & Remote Sensing*, 68(7), pp. 725-734.

ACKNOWLEDGEMENTS

This project was partially funded by the Council on Research and Creativity of the Florida State University, China's National Key Laboratory of Urban and Regional Ecology, and the International Geographic Information Fund of the American Association of Geographers.