PERGAMON

AVAILABLE AT
www.ComputerScienceWeb.com
POWERED BY SCIENCE DIRECT

Neural Networks 16 (2003) 547–553

Neural Networks

www.elsevier.com/locate/neunet

2003 Special issue

Application of four-layer neural network on information extraction

Min Han*, Lei Cheng, Hua Meng

School of Electronic and Information Engineering, Dalian University of Technology, Dalian 116023, China

Abstract

This paper applies neural network to extract marsh information. An adaptive back-propagation algorithm based on a robust error function is introduced to build a four-layer neural network, and it is used to classify Thematic Mapper (TM) image of Zhalong Wetland in China and then extract marsh information. Comparing marsh information extraction results of the four-layer neural network with three-layer neural network and the maximum likelihood classifier, conclusion can be drawn as follows: the structure of the four-layer neural network and the adaptive back-propagation algorithm based on the robust error function is effective to extract marsh information. The four-layer neural network adopted in this paper succeeded in building the complex model of TM image, and it avoided the problem of great storage of remotely sensed data, and the adaptive back-propagation algorithm speeded up the descending of error. Above all, the four-layer neural network is superior to the three-layer neural network and the maximum likelihood classifier in the accuracy of the total classification and marsh information extraction.

Keywords: Remote sensing; Neural network; Information extraction; Marsh

1. Introduction

Remotely sensed images distinguish different landform mainly through the different brightness of pixels. Different brightness of pixels shows the different information of spectrum. According to this, classifying the land cover to categories and then extracting needed category is an important application of the remotely sensed images. Traditionally, the ways to classify the land cover are based on the Bayesian classifier. It is well known that the Bayesian classifier is theoretically optimal if the assumptions about the probability density functions (PDFs) are correct. Poor performance may be obtained if the true PDFs are different from those assumed by the model (Heermann & Khazenie, 1992). The need to have a specific probabilistic model is a major limitation of the Bayesian approach. Maximum likelihood classification is based on the Bayesian classifier. It has the minimum value of classification error probability. We assume all kinds of targets on the ground submit to normal distribution when we conduct the statistic classification of image. According to this assumption, a distribution of a response pattern can be completely presented by the mean vector and variance. The statistic probability of some pixel belonging to each category can be computed through these parameters.

Artificial neural network has been greatly applied in the field of classification and information extraction of remotely sensed image, especially back-propagation neural network classifier. Back-propagation neural network has succeeded in realizing the supervised classification of remotely sensed image (Kanellopoulos & Wilkinson, 1997; Li, 1998). Yang and Zhou (2001) presented a classification method based on knowledge. Compared with traditional method, neural network does not need the probability pattern but becomes a classifier by its adaptive learning ability. It starts with an inherently parallel processing, adaptive learning and nonlinear technique. The problems solved by neural network usually focus on the small data sets less than 200 patterns. However, remotely sensed image has great training data sets usually more than 1000. Therefore, we should select the network structure and learning algorithm carefully when we use neural network to classify the remotely sensed image and extract needed category, so that neural network can solve the problems more efficiently and precisely.

Multilayer feed-forward neural networks realize the complex nonlinear mapping. Single-hidden-layer structure can solve most classification and information extraction problems, but if samples are comparatively complex, that is,

* Corresponding author.
E-mail address: minhan@dlut.edu.cn (M. Han).

0893-6080/03/$ - see front matter © 2003 Elsevier Science Ltd. All rights reserved.
doi:10.1016/S0893-6080(03)00120-5

they are highly in discretization, two-hidden-layer neural network can be adopted (Zhang, 1999). The general back-propagation has some disadvantages, such as long-time training and difficulty of converge, etc. Bruzzone and Serpico (1997) presented a technique—training for two steps—to accelerate the training time. Du, Mei, and Li (1998) introduced conjugate gradient with line search (CGL) to optimize the learning rate. In this paper, the model of remotely sensed image is built with a four-layer feed-forward neural network, and an adaptive back-propagation algorithm based on a robust error function is introduced to avoid the problem of long-time training and difficulty of convergence. Comparing the four-layer neural network with the three-layer neural network and the maximum likelihood method through the analysis of classification process and marsh information extraction results, conclusion can be drawn that the four-layer neural network is able to perform best in the accuracy of the marsh information extraction.

2. Algorithm description of BP neural network

Land cover classification of remotely sensed image should be done before marsh information extraction. The main idea of remotely sensed image classification using neural network is to make the feature of the image as the input signals, to train the neural network following some rule, then to classify output signals. In this paper, a four-layer neural network with two-hidden-layer is used to build the model of remotely sensed data. In Fig. 1 the structure of the four-layer neural network is shown.

In all kinds of algorithms of neural network, the back-propagation algorithm has gained nearly full development. However, it still has several limitations, for example, it easily falls into local minimum, cannot acquire the global optimum solution, its convergence rate is slow and so on. It is difficult to determine the learning rate, but the learning

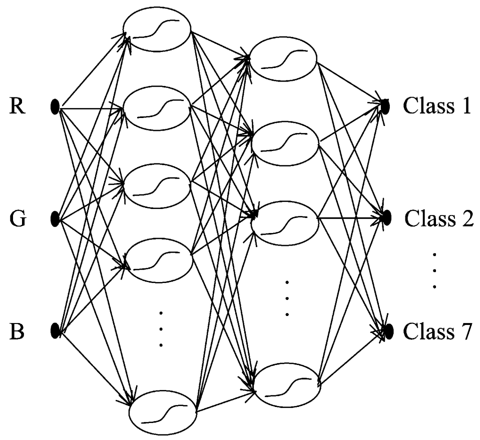

Fig. 1. Structure of four-layer feed-forward neural network.

rate is the key to influence the efficiency of the back-propagation learning algorithm (Wang, Li, & Mao, 2001; Zhou, Luo, Yang, Yang, & Liu, 2001). In normal back-propagation learning algorithm, the value of learning rate is constant. However, the optimum learning rate should be adjusted on line; training process is done under the direction of summed squared error function and reaches the minimum through the weight adjustment. This kind of error function is sensitive to the great error (Xia & Bian, 1998). In this paper, a robust error function is adopted to train the neural network, and an adaptive back-propagation algorithm, which is later proved to suit to classify the Thematic Mapper (TM) image of Zhalong wetland in China, is introduced to adjust the learning rate on line. The principle of the learning algorithm is introduced as following.

Firstly, define the new robust error function $\varphi(\varepsilon)$ under the following conditions ($\varepsilon = d_j(k) - y_j(k)$ is called the remain error):

(a) $\varphi(\varepsilon) = 0$, when $\varepsilon = 0$;
(b) $\varphi(\varepsilon)$ is continuous everywhere;
(c) $\varphi(\varepsilon)$ is symmetrical;
(d) $\Phi(\varepsilon) = (\partial\varphi(\varepsilon)/\partial\varepsilon) \geq 0$, when $\varepsilon \geq 0$;
(e) $\lim_{\varepsilon\to\infty} \Phi(\varepsilon) = 0$.

Typically, $\varphi(\varepsilon)$ can be selected as:

$$\varphi(\varepsilon) = \frac{\varepsilon^2}{b + a\varepsilon^2} \tag{1}$$

Then

$$\Phi(\varepsilon) = \frac{\partial\varphi(\varepsilon)}{\partial\varepsilon} = \frac{2b\varepsilon}{(b + a\varepsilon^2)^2} \tag{2}$$

The parameter a, b in equation should be selected according to the actual need. In this paper, we select $a = 1, b = 2$.

From the above equation, we can see that $\Phi(\varepsilon)$ is changing normally when ε is small. However, the increasing speed of $\Phi(\varepsilon)$ drops when ε is bigger, this leads to depress the influence of great error. Therefore, $\varphi(\varepsilon)$ is a robust function. When $a = 0, b = 1$, we have $\Phi(\varepsilon) = \varepsilon^2$, which is the error energy function in normal back-propagation learning algorithm.

Suppose the number of samples is P. Input a sample whose serial number is k, then the algorithm flow can be described as follows.

Step 1: Forward-propagation process. The relation is

$$y_j^l(k) = f\left(\sum_i \omega_{ji}^l(k) y_i^{l-1}(k) - \theta_j^l\right) \tag{3}$$

where $y_j^l(k)$ is the output of the jth unit of the lth layer. $\omega_{ji}^l(k)$ is the weight connecting ith unit of the $(l-1)$th layer and jth unit of the lth layer. θ_j^l is threshold value of the jth unit of the lth layer. $f(\cdot)$ is the active function. The active function of the hidden layer is nonlinear function usually, such as a sigmoid function and a hyperbolic function.

Step 2: Back-propagation process. It is also the adjustment process of the weight. The relation is

$$\omega_{ji}^l(k+1) = \omega_{ji}^l(k) + \mu(k)\Delta\omega_{ji}^l(k) \tag{4}$$

where μ is the adaptive learning rate, $\Delta\omega_{ji}^l(k)$ is the adjusted value of the weight. It can be carried out through the gradient descent algorithm.

$$\Delta\omega_{ji}^l(k) = \delta_j^l(k)y_i^{l-1}(k) \tag{5}$$

For output units, we have

$$\delta_j(k+1) = y_j(k)[1 - y_j(k)]\Phi(\varepsilon_j) \tag{6}$$

For hidden units, we have

$$\delta_j^l(k+1) = y_j^l(k)[1 - y_j^l(k)]\sum_q \delta_q^{l+1}(k)\omega_{qj}^{l+1}(k) \tag{7}$$

Train the neural network by inputting the neural network input–output samples. When P samples have been input, compute the total squared error E, which can be defined as

$$E = \frac{1}{2P}\sum_{k=1}^{P}\sum_{j=1}^{m}\varphi(\varepsilon_j(k)) \tag{8}$$

Step 3: Input P samples again and compare the total error acquired in two epochs. Suppose k is the serial number of circulation. The comparative results could be divided into three situation:

1. If $E_k < E_{k-1}$, that is the error is descending, learning direction is right, learning rate μ can be increased as follows:

$$\mu_{k+1}^l = \beta\mu_k, \qquad \beta \geq 1 \tag{9}$$

If $\mu_{k+1}^l > \mu_{\max}$, $\mu_{k+1} = \mu_{\max}$, otherwise $\mu_{k+1} = \mu_{k+1}^l$.

2. If $E_k > E_{k-1}$, it shows that learning step length is too long, then

$$\mu_{k+1}^l = \alpha\mu_k, \qquad \alpha < 1 \tag{10}$$

If $\mu_{k+1}^l < \mu_{\min}$, $\mu_{k+1} = \mu_{\min}$, otherwise $\mu_{k+1} = \mu_{k+1}^l$.

3. If $E_k > \lambda E_{k-1}$ $(\lambda > 1)$, make kth learning result unused, decrease the learning rate and train the neural network again.

In a word, TM image model of Zhalong wetland is established by four-layer feed-forward neural network; training process is done under direction of a robust error function and weight value is adjusted by the adaptive back-propagation algorithm.

3. Marsh information extraction of Zhalong wetland using neural network

In this paper, a four-layer feed-forward neural network is adopted to classify the TM image model of Zhalong wetland and then extract marsh information. The information extraction results are compared with three-layer feed-forward neural network and the maximum likelihood classifier.

3.1. Research area and its data

Our research area, that is named Zhalong protected district, lies in the lower reaches of Wuyuer river in the west of Heilongjiang province in China. Its geographic coordinate is 46°52′–47°32′N and 123°47′–124°37′E. Zhalong protected district is an upcountry wetland and ecosystem which is mostly covered by reed marsh, and it is the most complete, original and broad wetland ecosystem in the north of China compared to area of the same latitude. It has been listed in the international important wetland contents. Therefore, it is important to monitor the marsh change. Marsh information extraction from the remotely sensed image is an effective way to acquire the situation of the marsh.

The data we used for land cover classification and information extraction is half of a Landsat TM scene of the surroundings of Zhalong wetland. Date of TM image of Zhalong wetland acquisition is October 5, 2001.

The TM scene contains seven bands. In this research, we use combination of TM band 4, 3 and 2 to create a false color composite as seen in Fig. 2, which is suitable for land

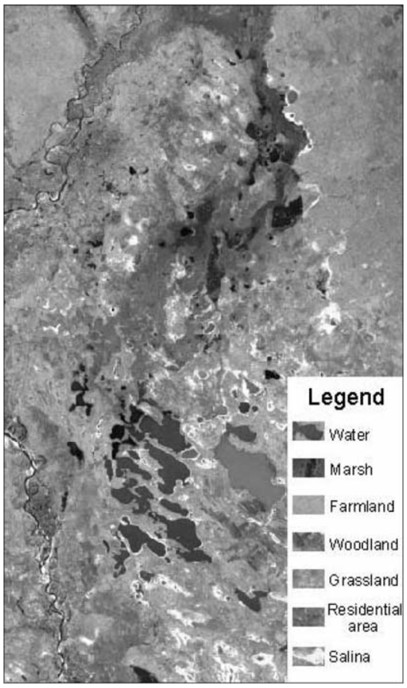

Fig. 2. Bands 4,3,2 composite image.

cover classification. The image was interpreted by comprehensive analysis and compared with map of land use and ability of visual interpreters. Then seven classes were confirmed in the region of TM image, which include water, marsh, farmland, woodland, grassland, residential area and salina.

3.2. Build the neural network model

Select the training and testing data. According to the results of visual interpreter, we select 1500 pixels from the false color composite to train the network and 700 pixels to test the network. The maximum likelihood classifier adopted the same samples. Table 1 shows the sample examples of every class in training sets and the numbers of every class samples, where TM4, TM3 and TM2, respectively, present the sample value of TM band 4, 3 and 2.

Before classification of remotely sensed image using neural network classifier, the samples should be changed into the region of $[0, 1]$. In this way, the supersaturated phenomenon could be avoided.

Confirm the parameter of input and output units of the network. The number of input units is three according to RGB color, and the number of output units is seven according to the number of categories. Each output unit presents a kind of category. The outputvector can be described as follows: (0.9,0.1,0.1,0.1,0.1,0.1,0.1),(0.1,0.9,0.1,0.1,0.1,0.1,0.1),… (0.1,0.1,0.1,0.1,0.1,0.1,0.9). An input pattern will be divided to ith category if it has the highest output value in ith unit.

The number of hidden layers and the number of hidden units are obtained by experiments. In the experiments, we found out that single hidden layer network structure cannot get the needed accuracy. When we adopted the two hidden layer network structure, we found out that the number of first hidden layer should be at least twice the number of input layer units. The number of second hidden layer should at least be the larger one of the input units and the output units. If we increase the number of the hidden layer units, accuracy can be improved, but it needs more learning time.

Parameters of the neural network can be obtained by such method. The result is, the number of the input units is three; the number of the hidden layer is two; the number of first hidden layer units is eight; the number of second hidden layer units is seven; the number of the output units is seven; the initial learning rate is 0.1; the initial weight and threshold value is random numbers. The structure of the neural network is shown in Fig. 1.

Compare the classification and extraction results of neural network and maximum likelihood classifier. The classification process of neural network is mainly realized by two steps. The first step is to train the network according to the samples; the second step is to classify the whole remotely sensed data using learning results.

Firstly, we should input the training data into the model of the neural network and train the neural network. The four-layer neural network adopts the adaptive back-propagation learning algorithm introduced in this paper and the three-layer neural network adopts normal back-propagation learning algorithm. The learning error curve of the three-layer neural network and the four-layer neural network are shown in Fig. 3, where error is computed by summed squared error function. We can see the error of four layer neural network falls faster than the three layer neural network, and the convergence error is also lowest than the three layer neural network. Therefore, the two hidden layer structure can reach more accurate degree compared with the single layer structure. At the same time we found out that the error function would not be convergent in simulation if we did not adopt the adaptive learning algorithm. It was noticeable that the robust error function makes the training time shorter comparing with normal summed squared error function.

The error matrix can be obtained from the classification results of test samples. The classification error matrices of four-layer neural network, three-layer neural network and maximum likelihood method are, respectively, shown in Tables 2–4. From the three tables, we can see that the four-layer neural network can get best classification results. Three-layer neural network mistook other classes of land cover as marsh and maximum likelihood classifier had weak ability to judge the grassland and residential area.

Table 1
Sample examples of training sets

Class	Band			Number
	TM4	TM3	TM2	
Water	18	63	93	400
Marsh	112	97	99	200
Farmland	195	190	162	200
Woodland	147	67	92	100
Grassland	73	90	106	300
Residential area	85	133	140	200
Salina	215	255	253	100

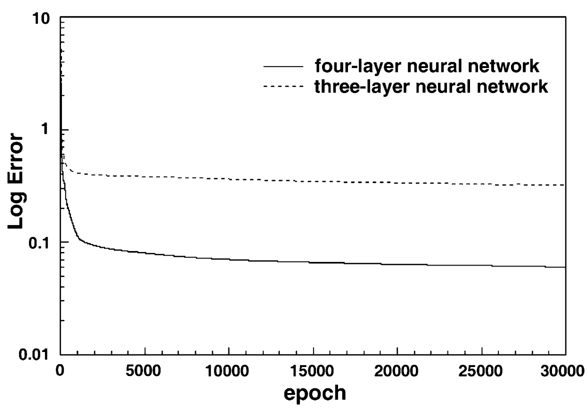

Fig. 3. Learning error curve of four-layer and three-layer neural network.

Table 2
The error matrix of four-layer neural network classification

Practice	Class							Total
	Water	Marsh	Farmland	Woodland	Grassland	Residential Area	Salina	
Water	99	0	0	0	0	2	0	101
Marsh	0	98	0	2	0	0	0	100
Farmland	0	1	100	2	0	0	0	103
Woodland	0	0	0	96	0	0	0	96
Grassland	0	1	0	0	93	3	0	97
Residential area	1	0	0	0	7	95	0	103
Salina	0	0	0	0	0	0	100	100
Total	100	100	100	100	100	100	100	700

Table 3
The error matrix of three-layer neural network classification

Practice	Class							Total
	Water	Marsh	Farmland	Woodland	Grassland	Residential Area	Salina	
Water	97	2	0	0	0	2	0	101
Marsh	1	83	0	0	7	0	0	91
Farmland	0	5	87	1	1	0	0	94
Woodland	0	0	0	94	0	0	0	94
Grassland	0	10	13	5	92	4	0	124
Residential area	2	0	0	0	0	94	0	96
Salina	0	0	0	0	0	0	100	100
Total	100	100	100	100	100	100	100	700

Classification accuracy calculations on pixels are based on the error matrix. Accuracy of each class can be calculated by the error matrix shown in Tables 2–4. As seen in Table 5, four-layer neural network can obtain highest accuracy of each class.

In most cases, the integrated classification accuracy reflects the total effect of classification. The integrated classification accuracy is average classifier accuracy. The integrated classification accuracy of four-layer neural network, three-layer neural network and maximum likelihood classifier is, respectively, 97.29, 92.43 and 89.43%. It shows that the classification result of four-layer neural network is better than three-layer neural network and maximum likelihood classifier in integrated accuracy, which presents that the four-layer neural network adopted in this paper is more suitable to build complicated model of TM image.

After the classification of the whole remotely sensed image using trained neural network and maximum likelihood classifier, the result images of marsh information

Table 4
The error matrix of maximum likelihood classification

Practice	Class							Total
	Water	Marsh	Farmland	Woodland	Grassland	Residential area	Salina	
Water	87	0	0	0	0	4	0	91
Marsh	6	95	4	2	22	0	0	129
Farmland	0	2	95	0	4	0	0	101
Woodland	0	0	0	96	0	0	0	96
Grassland	0	3	1	2	66	9	0	81
Residential area	7	0	0	0	8	87	0	102
Salina	0	0	0	0	0	0	100	100
Total	100	100	100	100	100	100	100	700

Table 5
Accuracy of the three methods

Method	Class							Average accuracy (%)
	Water (%)	Marsh (%)	Farmland (%)	Woodland (%)	Grassland (%)	Residential area (%)	Salina (%)	
Four-layer neural network	99.00	98.00	100.00	96.00	93.00	95.00	100.00	97.29
Three-layer neural network	97.00	83.00	87.00	94.00	92.00	94.00	100.00	92.43
Maximum likelihood	87.00	95.00	95.00	96.00	66.00	87.00	100.00	89.43

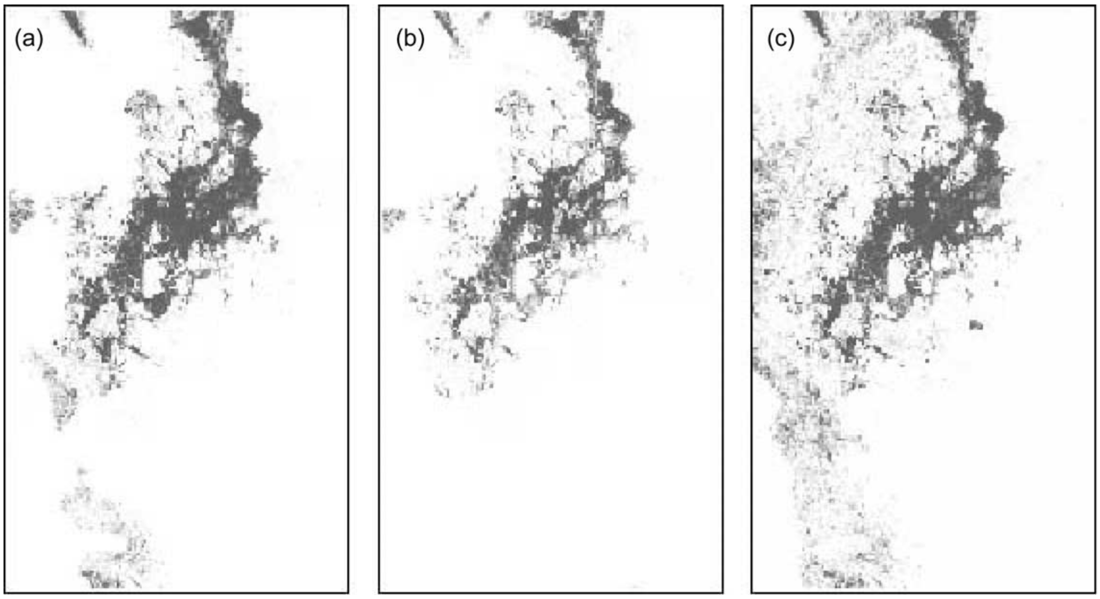

Fig. 4. Marsh information extraction results with three methods: Marsh information extraction with (a) four-layer neural network; (b) three-layer neural network; (c) maximum likelihood classifier.

extraction are obtained and shown in Fig. 4(a)–(c). Since the four-layer neural network can get highest accuracy of each category (including marsh), the accuracy of marsh information extraction is highest.

4. Conclusion

The four-layer neural network and the adaptive back-propagation algorithm based on the robust error function appear to be feasible to classify the TM image of Zhalong wetland and extract marsh information. Compared to three-layer neural network, four-layer neural network could avoid the burden of large storage; Compared to the traditional maximum likelihood classifier, neural network needs not to assume the model of probability but acquires weight value through learning ability then becomes a classifier. Therefore, it avoided the error made by the imprecise probability model. From the simulation results, mistaken and lack classification phenomena of four-layer neural network decreased greatly. Therefore, the classification and information extraction accuracy of four-layer neural network is superior to three-layer neural network and the maximum likelihood classifier; the adaptive back-propagation learning algorithm based on the error robust function avoids the divergence of error function, decreases the general error to a smaller degree and accelerates the learning rate. These lead the marsh information extraction to be conducted successfully. Other neural network structure and algorithm should be taken to try to improve the information extraction accuracy.

Acknowledgements

This research is supported by the project (50139020) of the National Natural Science Foundation of China. All of these supports are appreciated.

References

Bruzzone, L., & Serpico, S. B. (1997). Classification of imbalanced remote-sensing data by neural networks. *Pattern Recognition Letters*, *18*, 1323–1328.

Du, H., Mei, W., & Li, D. (1998). Classification of remote sensing images using BP neural network with dynamic learning rate. *Journal of Beijing Institute of Technology*, *18*(4), 485–488.

Heermann, P. D., & Khazenie, N. (1992). Classification of multispectral remote sensing data using a back-propagation neural network. *IEEE Transactions on Geoscience and Remote Sensing*, *30*(1), 81–88.

Kanellopoulos, I., & Wilkinson, G. (1997). Strategies and best practice for neural network image classification. *Remote Sensing*, *18*(4), 711–725.

Li, Z. (1998). Supervised classification of multispectral remote sensing image using B-P neural network. *Journal of Infrared and Millimeter Waves*, *17*(2), 153–156.

Wang, Y., Li, S., & Mao, J. (2001). *Image processing and recognition with computer*, (pp. 197–205). Beijing: Higher Education Press.

Xia, G., & Bian, X. (1998). BP algorithm based on robust error function. *Journal of Harbin Engineering University*, *19*(3), 33–37.

Yang, C., & Zhou, C. (2001). Investigation on classification of remote sensing image on the basis of knowledge. *Geography and Territorial Research*, *17*(1), 72–77.

Zhang, J. (1999). Construction of a neural network on remote sensing image classification. *Journal of Image and Graphics*, *4(A)*(10), 831–834.

Zhou, C., Luo, J., Yang, X., Yang, C., & Liu, Q. (2001). *Geography understanding and analysis of remotely sensed image*, (pp. 229–239) Beijing: Science Press.