

A INTEGRATED CLASSIFIER FOR LARGE REGIONAL-SCALE LAND-COVER CLASSIFYING

Chengfeng LUO Zhengjun Liu Wenli Meng

Key Laboratory of Mapping from Space of State Bureau of Surveying and Mapping, Chinese Academy of Surveying and Mapping, Beijing 100039, China - (chfluo, zjliu, mengwl@casm.ac.cn

Commission VII, WG VII/4

KEY WORDS: Remote sensing, Land Cover, Classification, Algorithms

ABSTRACT:

With a lot of successful applications of neural network-based classification, it has been recognized the classification can produce more accurate results than conventional approaches for remotely-sensed data. Although its training procedure is sensitive to the choice of initial network parameters and to over-fitting, the multilayer feed-forward networks trained by the back-propagation algorithm is adopted by many study cases. To overcome the shortages the genetic algorithm is involved, for one has the advantage during searching the infinitesimal point in local space and the other in global space. The integrated method can avoid the risk of premature convergence while BP network was training, moreover the end total mean square error of the network is steadier even if you redo the whole course several times, and turn the instability of classification result away. In the article, the algorithm mentioned above was used to work Chinese surface land cover classification at coarse spatial scales with input parameter dataset from MODIS data, at the same time two other algorithms: Fuzzy ARTMAP ANN and Maximum likelihood were involved for a comparison purpose. The result shows that the error back propagation neural network and genetic algorithm integrated method not only run with better efficiency, but also produce the best result than other two classifiers.

1. INTRODUCTION

Artificial Neural Networks (ANN) have been widely used for information extraction in remote sensing in the last 20 years. ANN algorithm relies on the data themselves and does not need the statistical hypothesis of data which is precondition for classifiers based statistic like Maximum likelihood classifier. For remote sensing data, it is difficult to extract the information implied in remote sensing data with routine statistic method because of the complex data distribution and illegibility and uncertainty of the data, which are induced by complicated radiation of the earth's surface objects and many dubious disturbances while imaging. Many studies have been done with ANN classifiers using different remotely-sensed datasets (Benediktsson et al., 1990,1993; Hara et al., 1994; Tzeng et al., 1994;).In nearly all cases, the neural network classifiers have proven superior to conventional classifiers, often recording overall accuracy improvements in the range of 10-20 percent (Strahler,A.,1999).

While there are a variety of different neural network models, most remote sensing practical applications have used a supervised, feed forward structure employing an error back propagation (BP) algorithm and its deformations. The gradient descent technique is the optimization tool for BP training and sigmoid function is selected as the transfer function, which would induce the training result to the local minimum, and unsteady training result, low execute and convergence speed.

In the article BP neural network classifier based on genetic algorithm was proposed to classify with remote sensing data at coarse spatial scales. The method makes up the shortcoming of BP algorithm during the classification process with remote sensing data. The method was used to make the land-cover classification experiment in whole china with MODIS data at

1km scale. Two other classification algorithms, fuzzy ARTMAP ANN and Maximum likelihood classifier, were selected to compare. The result shows that the classification map with proposed method preceded the other classification results obviously. The execute efficiency of the method proposed in the article is far better than that of fuzzy ARTMAP classifier.

2. METHODOLOGY

Error Back Propagation algorithm was proposed by Rumelhart in 1985. The whole network learning process comprises forward and backward parts. The input variables are fed forward through the network to produce an output vector. During a following back-propagation phase, the synapse weights are adjusted so that the network output vector more closely matches the desired output vector. The network weights, or processing element responses, are adjusted by feeding the summed squared errors from the output layer back through the hidden layers to the input layer. The network cycles through the training set until the synapse weights have been adjusted so that the network output has converged, to an acceptable level, with the desired output. The basic equations relevant to the back-propagation model are presented in Fischer and Gopal (1992).

Genetic Algorithm (GA) is an optimization algorithm to search random which is based on natural selection and genetic mechanism in biology (S. Bandyopadhyay, 2001) . The essential characteristics of GA are colony policy and simple genetic operators, which breaches the local search limit and can search, gather and inherit information in the whole solution space. Unlike the case in gradient descent-based training algorithms, the fitness (or error) function does not have to be differentiable or even continuous since GA does not depend on

gradient information. So it can handle large, complex, nondifferentiable and multimodal spaces, which are the typical cases in remote sensing classification and many other real world applications.

The combination of GA and BP for weight training consists of three major phases. Firstly, build the representation of connection weights, that is decode each individual (genotype) in the current generation into a set of connection weight; secondly, construct the fitness function which is used to evaluate the fitness of an individual; thirdly, decide the evolutionary policy such as initial colony, selection, crossover and mutation operations by a GA according to its fitness.

The typical evolutionary process for BP connection weights can be described as:

Decode each individual (genotype) in the current generation into a set of connection weights. In the article the real number form was adopted to represent the connection weights and each neuron's connection weights and bias was set to its correspondent gene segments.

Evaluate each set of the connection weights by construction the corresponding neural network structure and computing its total mean square error between actual and target outputs. The fitness of an individual is determined by the total mean square error (MSE). The higher the error, the lower the fitness, the fitness function is defined as:

$$\varepsilon (net) = \sqrt{\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^k (T_i^j - f_i^j)^2} \quad (1)$$

Where ε is MSE, net is the BP network, f_i^j is the actual neuron output for the input training sample, while T_i^j its desired response.

Select parents of reproduction based on their fitness. Roulette wheel selection scheme is adopted in the experiment (D.E. Goldberg.,1989; J.H. Holland.,1975) .the population of current generation is mapped onto a roulette wheel, where each chromosome is represented by a space that proportionally corresponds to its fitness.

Apply search operator in conjunction with the crossover and/or mutation operators, to parent chromosomes to generate offspring, which form the next generation. An asexual reproduction operator, arithmetical crossover operator and a non-uniform mutation operator are applied in the experiment of the article.

The method mentioned above take the advantage of the GA to search the near-optimal finest solution globally, which uses many Genetic operators to find the finest connection weights in the whole solution space without computing gradient information and initiating the network. The execute efficiency can be improved greatly, but some time the solution is not the best which is not our final object. So the search was continued with gradient descent training based near-optimal connection weights until the best-optimal connection weights were met.

BP and GA Integrated Method avoid the BP's local convergence. Moreover gradient descent training was only used to search the local optimal connection weights after the GA, which induce not only fast execute speed but also steady initial connection weights for BP training. The final connection weights and classification result are steady-going even if the training were repeated sever times.

3. DATA

Our remote sensing data come from Land Process Distributed Active Archive Centre, U.S.A, which include MODIS products MOD13A1 and MOD11A2 at 1-km spatial resolution in Chinese area. Geometry correction software, MRT developed by USGS EROS data centre, was employed to correct and beset the data.

The monthly Ts and NDVI were produced with the maximum value composite from the MODIS products, and we simulated the lost Ts because of cloud pollution; two indexes Temperature-Vegetation Angel (TVA) and Temperature-Vegetation Distance (TVD) were constructed based NDVI/Ts Space (Lambin, E. F, 1996); then monthly TVA and TVD were computed. Primary Components Analysis (PCA) was used to compress useful information into several principal components. Thus, lesser components replace former multidimensional data on the premise of losing least information, and the time for dealing with data can be reduced greatly; in addition PCA can remove the noise brought by poor data quality partly (Townshend, J. R,1987).

$$TVA = \frac{\arctg[Ts / (100 * NDVI)]}{\pi / 2} \quad (2)$$

$$TVD = \sqrt{Ts^2 + (100 * NDVI)^2} \quad (3)$$

After the PCA, the variances of the first three principal components for TVA are: 41.73%、26.72%、8.12%, accounts for 76.57% in whole information; For TVD, the variances of the first two principal components account for 89.46% in whole information, they are 78.89% and 10.57%. Without losing the main information only the dominating principal components were adopted as input parameters for classification in order to put down the time of classifying.

China is a country with wide-ranging territory and diversiform, complicated hypsography. Most regions are mountainous in west; there are altiplano and champaign in Middle part; the topography in East is low and plain except the big xin-an mountain range. Thus we selected spatial structure and digital elevation model information as classification for improving classification accuracy.

The spatial structure information was described by spatial structure index. The TVA in December and TVD in June were used to compute the index because they had max information grads out of all time series data. Involving the five principal components and digital elevation model information, so far eight input parameters aggregately were got for classification.

4. THE CLASSIFICATION UNITS OF LAND COVER

The main classification feature was from NDVI & Ts, in view of remote sensing data and matter-of-fact instance in China, the classification units of land cover adopted include 17 sub-classes, which referenced the Vegetation Map of China at the scale of 1:4000000, IGBP system of units (Belward, A.S., et al, 1996) and other Chinese systems of units.

To select training signatures of various land-cover types for our study, we referenced (1) the Vegetation Map of China at the scale of 1:4000000, which was produced by the Institute of Botany, CAS; and (2) Land-use Map of China at 1-km spatial resolution in 2000, which was interpreted from TM images by Remote Sensing Application Centre for Resources and

Environment, IRSA, CAS; and (3) Land-cover Map of China at 1-km spatial resolution from MODIS images by Globe Change Research Group, IRSA, CAS, in 2001; and (4) check samples based on field surveys from 1990 to 1995 which was used primarily to define training signatures of forest types; and (5) TM images in part regions; and (6) literature and investigate data on the spot. 4370 samples, belonging to 17 types, were picked out, and 2952 points were used for training, others for testing classification accuracy. Jeffries—Matusita distance was used to measure the separability between land cover classes(the value ranges form zero to two, and it will be a good separability if the value is bigger than 1.8), among our samples the minimum separability is 1.94 between shrubs and Single maturity croplands.

Code	Land-cover types	Description
1	Evergreen broadleaf forest	Natural or planted broadleaf trees with a percent cover >60% and height exceeding 2 meters, which remain green all year
2	Evergreen needleleaf forest	Natural or planted needleleaf trees with a percent cover >60% and height exceeding 2 meters, which remain green all year
3	Deciduous broadleaf forest	Natural or planted broadleaf trees with a percent cover >60% and height exceeding 2 meters, which have an annual cycle of leaf-on and leaf-off periods
4	Deciduous needleleaf forest	Natural or planted needleleaf trees with a percent cover >60% and height exceeding 2 meters, which have an annual cycle of leaf-on and leaf-off periods
5	Mixed forest	Natural or planted mixed (broadleaf and needleleaf) trees with a percent cover >60% and height exceeding 2 meters
6	Typical grasslands	Lands dominated by dryland herbage plant with a percent cover between 20-40%
7	Hungriness grasslands	Lands dominated by drought-resistant herbage plant with a percent cover between 5-20%
8	Meadow grasslands	Lands dominated by all kinds of Herbaceous plant with a percent cover between 40-75%
9	shrubs	Closed shrubs and dwarf trees with a percent cover >60% and height less than 2 meters
10	Double maturities croplands	Lands covers with dryland crops that are double in one year or three maturities in two years
11	Single maturity croplands	Lands covers with temporary crops followed by harvest and a bare soil period; and drought-resistant economical crops
12	Marsh	lands with a permanent mixture of water and herbaceous of woody vegetated that cover extensive areas
13	Hungriness and barren	Lands with hungriness, gobi, exposed rocks etc., which never has more than 5% vegetated cover during any time of the year.
14	Sand desert	Sand desert and running sand-dune
15	City and Built-up lands	Land covered by buildings and other man-made structures.
16	Water Bodies	Lakes, reservoirs, and rivers on land, which can be either fresh or salt-water bodies
17	Ice and snow	Lands under snow or ice cover throughout the year

Tab.1 Description of the remotely sensed land-cover system for China

5. RESULT

For the BP and GA neural network structure, a three layer with one hidden layer consisting of thirty hidden nodes was used (Schowengert, 1995). Input nodes were set eight (8 bands), and output nodes seventeen (17 land cover classes). The BP network connection weights and biases was decoded chromosome orderly which had 797 genes ranged from -2 to 2.

For GA process, the initial population is 160, and asexual reproduction probability was set 0.8, arithmetical crossover probability 0.09, and mutation probability 0.11. The GA was trained by 60 generation, followed by a BP training procedure whose learning step was set 0.1, increased by 1.1, and objective

error 0.001, iterative time 4000. The final actual network error is 0.00537.

With the input parameters and classifier mentioned above, Land cover classification map for china in 2004 was produced showed as fig 1.

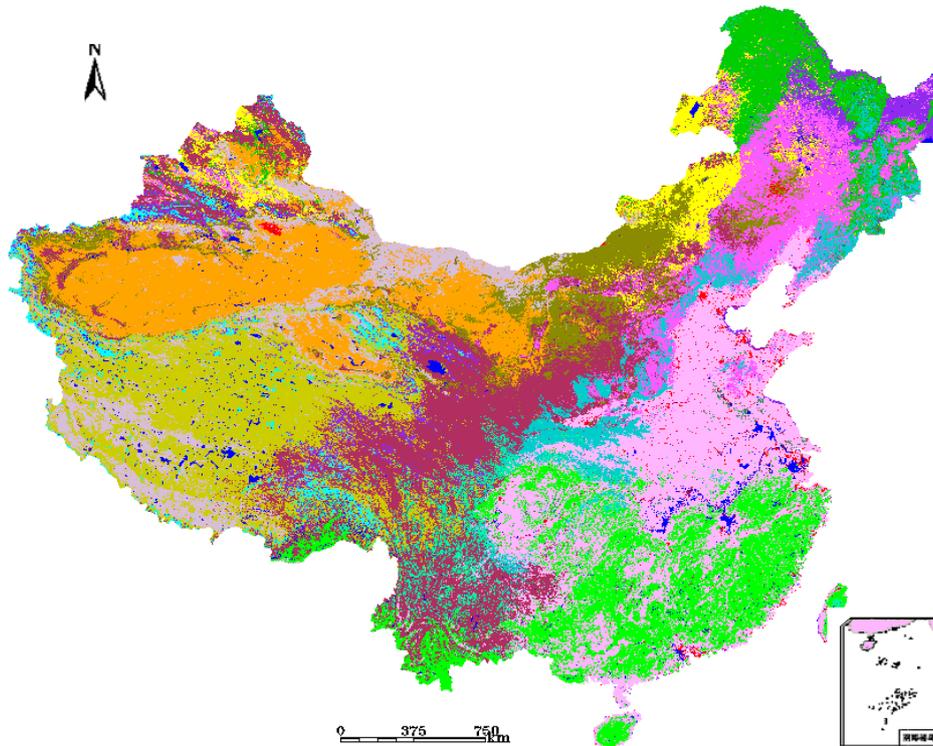
BP neural network classifier based on genetic algorithm, fuzzy ARTMAP ANN (Gail A.,1992) and Maximum likelihood classifier were compared for their performance to produce Chinese surface land cover classification result with remotely-sensed data at coarse spatial scales. The compare based on the classification map, area of different classes and accuracy evaluation. With the same input parameters and samples, the

main difference of classification results can be showed as following.

- (1) The accuracy evaluation with confusion matrix shows that the accuracy of BP neural network classifier based on genetic algorithm is the best, which total accuracy is 89%, Kappa coefficient 0.89. And the total accuracy of fuzzy ARTMAP classifier is only 39%. For producer and user accuracy of most single class, fuzzy ARTMAP classifier and Maximum likelihood classifier are far lower than BP neural network classifier.
- (2) There is an obvious error in the classification result with Maximum likelihood classifier. The total water area is 11.8% and the region around desert in north Sinkiang was mistook for water. The City and Built-up lands in north sienna altiplano is another obvious mistake. The inaccuracy comes out from the

algorithm itself because the separability between water and other classes is more than 1.98.

- (3) The classification result of fuzzy ARTMAP classifier is unsuccessful. In classification map, most area is covered by shrubs, hungriness grasslands and marsh, which is an obvious mistake. Almost all classes' spacial location and area proportion disaccord with the fact. The result is even worse than that of Maximum likelihood classifier.
- (4) Compared with the Vegetation Map of China at the scale of 1:4000000, in the classification map from BP neural network classifier based on genetic algorithm, most classes' spacial geographic distribution are accordant with the vegetation map. The forest area proportion is 19.4%, and croplands 21.2%, shrubs 14.1%. And the proportion of croplands is on the high side, while that of shrubs is on the low side.



Legend

 Evergreen broadleaf forest	 Deciduous needleleaf forest	 Mixed forest
 Evergreen needleleaf forest	 Deciduous broadleaf forest	 Double maturities croplands
 Typical grasslands	 Meadow grasslands	 Hungriness grasslands
 Single maturity croplands	 Shrubs	 City and Built-up lands
 Hungriness and barren	 Marsh	 Sand desert
 Ice and snow	 Water Bodies	

Figure.1 Land cover classification map for China

6. CONCLUSION

- (1) The MLC classifier based on statistic mistook non-water object for water and made the water area proportion 11.8% which is far different from the classification result with BP

neural network based on genetic algorithm classifier. With same samples and high separability, this situation proved that MLC can not distinguish some classes. MLC is based on statistic rules on the assumption that the remote-sensed data are

normal distribution. In fact, it is very difficult to meet the assumption because of uncertain remote-sensed data induced by many disturbance factors. On the other hand the MODIS data at coarse spatial scales implied mixed pixels, which raise the uncertain character, also affected the result.

(2) The result with fuzzy ARTMAP classifier should be a good one according as the basic theory and successful experiment with high resolution remote-sensed data, but in fact our result is too bad. We found that the main reason is parameter setting after many experiments. During the training process, the learning rate and vigilance parameter are key

factors determining the node number of F_2^a , which decides network structure. At the same time there is no orderliness to follow about how to set these parameters. While fuzzy ARTMAP classifier was used to work Chinese surface land cover classification, there is so many classes, samples and nodes in F_2^a that make the training and classifying speed very slow. In our experiments, the fuzzy ARTMAP classifier runs with far lower efficiency than BP neural network based on genetic algorithm classifier.

CODE	MLC		ARTMAP		BP&GA	
	user accuracy (%)	area proportion (%)	user accuracy (%)	area proportion (%)	user accuracy (%)	area proportion (%)
1	82.67	9.1	100.00	2.9	89.61	9.6
2	85.90	0.8	91.67	0.3	86.75	2.0
3	65.26	1.9	62.50	0.4	80.46	3.5
4	91.04	3.3	50.49	3.6	93.42	4.0
5	86.76	0.3	93.94	0.1	84.15	0.3
6	86.49	2.9	0.00	0.1	94.94	3.2
7	100.00	4.5	50.00	0.3	91.95	12.3
8	95.65	4.4	20.80	24.9	86.21	5.4
9	74.71	12.0	40.11	20.1	93.98	14.1
10	69.00	16.5	21.93	13.6	92.94	14.8
11	64.91	8.1	6.48	7.8	82.14	6.4
12	90.63	5.1	32.33	15.2	85.90	3.4
13	76.71	7.4	100.00	2.0	90.24	6.7
14	85.71	9.9	32.65	5.0	88.51	10.5
15	80.56	1.4	0.00	0.2	100.00	0.7
16	38.20	11.8	73.68	1.5	89.01	1.4
17	100.00	0.7	95.00	2.0	100.00	1.6

Table.2 Comparison of different classification results

	MLC	ARTMAP	BP&GA
Total accuracy coefficient	75.9684%	39.1679%	89.8135%
Kappa coefficient	0.7448	0.3537	0.8918

Table.3 Comparison of classification total accuracy of each algorithm

(3) The Error Back Propagation algorithm optimized by Genetic Algorithm gives attention to two optimization algorithm, genetic algorithm and error back propagation neural network algorithm, which have the advantage during searching the infinitesimal point in local space and global space respectively, avoiding the risk of premature convergence while BP network was training. Compared with BP algorithm, the end total mean square error of the network is steadier even if you redo the whole course several times, and turn the instability of classification result away. High classification accuracy and fast execute speed can be achieved when classifying with the method mentioned above using remotely-sensed data at coarse spatial scales.

(4) Chinese surface land cover classification result with the proposed algorithm in our experiment is not very ideal. Maybe one reason is the difference between simulative Ts dataset and the actual dataset. The other reason is some samples selected

maybe not be representative. Resource data with high quality and representative samples are foundational to get Chinese land cover classification map with high reliability.

REFERENCES

Belward, A.S., et al,1996.The IGBP-DIS globe 1 km land cover data set (DISCover)-proposal and implementation plans: IGBP-DIS Working Paper No.13, Toulours, France [R].

Benediktsson, J. A., P. H. Swain and O. K. Ersoy,1993. Conjugate-gradient neural networks in classification of multisource and very-high-dimensional remote sensing data [J]. *International Journal of Remote Sensing*, 14,pp: 2883~2903.

Benediktsson, J. A., P. H. Swain and O. K. Ersoy,1990. Neural network approaches versus statistical methods in classification of multisource remote sensing data [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 28 (4) ,pp: 540~552.

D.E. Goldberg. Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley New York, 1989.

Fischer, M. M. and S. Gopal,1992. Neural network models and interregional telephone traffic: Comparative performance between multilayer feedforward networks and the conventional spatial interaction model. WSG Discussion Paper 27/92,

Department of Economic and Social Geography, Vienna University of Economics and Business Administration.

Fitzgerald R W, Lee B G,1994. Assessing the classification Accuracy of multisource Remote Sensing data [J]. *Remote sensing of Environment*, 47,pp: 362-368.

Gail A. Carpenter, Stephen Grossberg, Natalya Markuzon, etc,1992. Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps [J]. *IEEE Transactions on Neural Networks*, 3, pp: 698-713.

Hara, Y., R. G. Atkins, S. H. Yueh, R. T. Shin, and J. A. Kong,1994. Application of neural networks to radar image classification [J]. *IEEE Transactions on Geoscience and Remote Sensing*,32: 100~111.

J.H. Holland. *Adaptation in Natural and Artificial Systems*. The University of Michigan Press, 1975.

Lambin, E. F. and D. Ehrlich,1996a. The surface temperature-vegetation index space for land cover and land-cover change analysis. *International Journal of Remote Sensing* 17(3):463-487.

Paola,J.D. and R.A.Schowengerdt,1995. A review and analysis of backpropagation neural networks for classification of remote-sensed multi-spectral imagery. *International Journal of Remote Sensing*, 1995a, 16(16),pp:3033-3058.

S. Bandyopadhyay, S.K. Pal,2001. Pixel classification using variable string genetic algorithms with chromosome differentiation, *IEEE Trans. Geosci. Remote Sens.* 39 (2) , pp: 303-308.

Strahler,A.,Muchoney,D.,Borak,J.,Feng,G.,Friedl,M.,Gopal,S.,Hodges,J.,Lambin,E.,McLver,D.,Moody,A.,Schaaf,C.,andWoodcock,C.,1999,MODIS land cover product algorithm theoretical basis document(ATBD),Version 5.0 [R].Boston University Center for Remote Sensing, Boston, MA.

Townshend, J. R., Justice, C. R., and Kalb, V,1987. Characterization and classification of South American land cover types using satellite data [J]. *International Journal of remote sensing*, 8: 1189-1207.

Tzeng, Y. C., K. S. Chen, W.-L. Kao and A. K. Fung,1994. A dynamic learning neural network for remote sensing applications [J]. *IEEE Transactions on Geoscience and Remote Sensing*,32, pp: 1096~1103.

ACKNOWLEDGEMENTS

This work was carried out under China's National Social Development Research Programme funded by the Ministry of Science and Technology (Project No: 2004DIB3J107), the support is greatly appreciated. Our colleagues are all acknowledged for their help and their contributions.