Paper presented at the 22nd Asian Conference on Remote Sensing, 5 - 9 November 2001, Singapore.

A KNOWLEDGE-INTEGRATED RBF NETWORK FOR REMOTE SENSING CLASSIFICATION

Jian-cheng LUO[1], Cheng-hu ZHOU[1], Yee LEUNG[2],

LREIS, Institute of Geographical sciences and natural resources research, CAS, Beijing, China[1]

Tel: (86)-10-64889694 Fax: (86)-10-64889630 Email: luojc@lreis.ac.cn

Department of Geography, The Chinese University of Hong Kong, Hong Kong[2]

KEY WORDS: Radial Basis Functions, Knowledge, Remotely Sensed Image, Classification

ABSTRACT: Most Artificial neural networks (ANN) models used in the remote sensing classification are based on the multilayer perceptron (MLP) with back-propagation (BP) training algorithm. Compared to conventional statistical classifiers, MLP classifiers are non-parametric and distribution-free and is thus less restrictive in approximation, especially when distributions of features are strongly non-Gaussian. However, the major shortcomings of this class of networks are that they take relatively longer time to train and are prone to convergence to local minimum. The radial basis function (RBF) network, which combines the characteristics of the parametric statistical distribution model and non-parametric single layer perceptron, train much faster and are more stable than BP while keeping similarly complicated proximity. ANN, however, are very efficient in performing in classification tasks with low level of intelligence. They are less capable of reasoning with deep level knowledge, which is generally symbolic in nature. To better approximate the reality, integration of ANN with symbolic geographical knowledge is thus essential in the remote sensing classification. A knowledge-integrated RBF model that combines the power of approximation in high-dimensional space of the RBF network and the logic reasoning of rule-based inference is proposed in the present study. In addition to conceptual and technical discussions of the model, our arguments are substantiated by a real-life application. The experimental results show that the proposed model is more accurate, faster in training, simple in structure, and more interpretable.

1. INTRODUCTION

Thematic mapping by the classification of satellite data via pattern recognition has been one of the most important methods in remote sensing. Artificial neural networks (ANN) have been extensively applied to perform information extraction and classification of remotely sensed data (Atkinson and Tatnall 1997). Compared to the symbolic systems, as a massive parallelly distributed architecture with a large number of units and connections, ANN can simulate the basic functions of human neural system, and it is especially suitable for the simulation of human vision. A great variety of ANN models have been proposed in the past several decades. The multilayer perceptron (MLP) with back-propagation (BP) algorithm might be one of the most widely used models for information extraction and classification. Compared to the conventional statistical classifiers, the BPNN is distribution-free and non-parametric, and is more robust, especially when the distributions of features are strongly non-Gaussian. However, BPNN exhibits some serious drawbacks such as slow convergence in learning phase, the potential convergence to local minimum, the common chaotic behavior, and the inability to detect over-fitting.

Radial basis function (RBF) network (Powell 1987), on other hand, is another type of multilayer network which is very different from BPNN in its training algorithm. In RBF network, the output units form a linear combination of the basis functions in the kernel layer, and the basis functions produce a localized response to the input. The basis function can be viewed as an activation function in the kernel layer in which each unit has a localized receptive field to the input vector. RBF networks can overcome some of the above limitations of BPNN by relying on a rapid training phase, avoiding chaotic behavior, having simpler architecture while keeping complicated mapping capability. Such characteristics and the intrinsic simplicity of the RBF networks make them an interesting alternative to pattern recognition in general (Bishop 1995, Bruzzone 1999).

One of the problems of the neural computation models is that they could only simulate low level cognitive functions of human vision and neural system resulting in an understanding of images with a very coarse degree (Medsker 1994). They are not effective in reasoning with deep knowledge that is generally represented in a symbolic way. In addition to spectral information, recognition and classification of remotely sensed images usually require domain specific knowledge such as DEM and its derivatives. To achieve more accurate classifications, neural networks and symbolic knowledge should be integrated into a single system (Foody 1995, Gong 1996, Peddle 1995, Murai and Omatu1997). Integrating geographical knowledge built on top of geographical information systems has become an approach to increase the accuracy and effectiveness in the classification of remotely sensed images. Such knowledge can be used to fine tune neural network classification based on spectral information.

In this paper, a knowledge-integrated RBF model for the remote sensing classification is proposed. The integrated model employs a RBF network to classify images with spectral information. Geographical knowledge represented as rules is parallelly used to classify the images with topographical information. Classification results obtained from both methods are then combined by some evidence combination methods to derive the ultimate classification of the images. The approach thus takes advantages of the best of both worlds. In section II, we first describe the overall architecture of the RBF model. In section III, the effectiveness of the RBF model is evaluated by an application. The paper is then concluded with a summary and outlook in section IV.

2. THE KNOWLEDGE-INTEGRATED RBF MODEL

2.1 The architecture of the knowledge-integrated RBF model

There are four major components in the knowledge-integrated RBF model for the remote sensing classification (Figure 1): (A) Data source, (B) RBF network, (C) rule-based inference, (D) evidence combination. The first component is data source management that processes and prepares remote sensing data for the neural-network classification and geographical information (from GIS) for the rule-based inference. The neural-network component is essentially a RBF network that performs land-cover classification by hyper surface reconstruction in high dimensional space. Embedded in the RBF network is the ART network which facilitates the learning phase by performing efficient clustering in the kernel layer of the RBF network. Parallel to the neural-network component is the rule-based inference engine which classifies land covers by topographical knowledge built on top of geographical information system. Classification results of the RBF network and the rule-based inference are integrated within the evidence-combination component to produce the final classification of land covers.

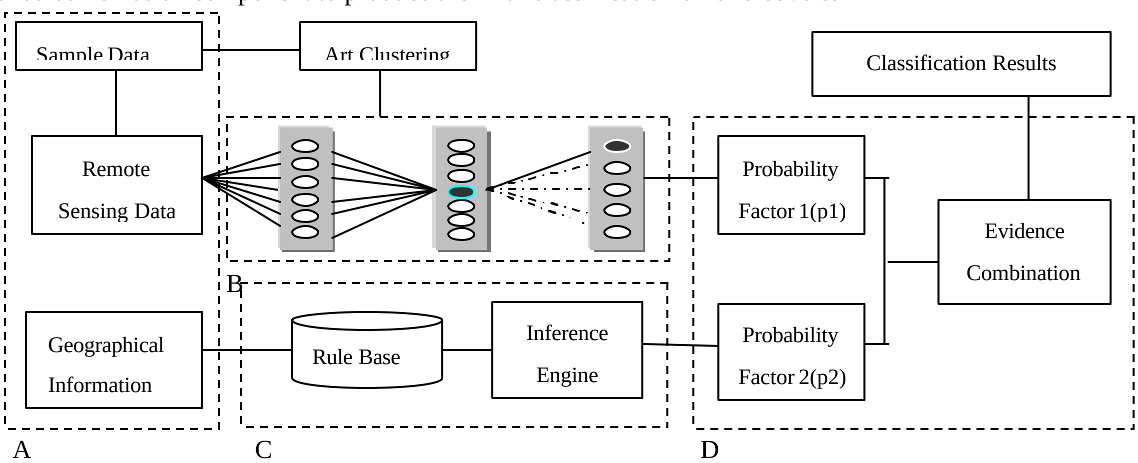

Figure 1. The general architecture of the knowledge-integrated model

(A) Data source; (B) RBF Network; (C) Rule-Based Inference (D) Evidence Combination

2.2 The RBF Network

RBF networks can be regarded as a special MLP structure in which the parametric statistical distribution model and non-parametric linear perceptron algorithm are combined together in serial sequence. The basic structure of a RBF network consists of an input layer (*I*), a kernel (hidden) layer (*K*), and an output layer (*O*). In the context of a ANN, the units in the kernel layer provide a set of kernel basis functions (called radial basis functions) that constitute the "basis" for the input vectors when they are expanded into the kernel unit space. The basis functions can be viewed as the activation functions in the kernel layer. The output of the RBF network is a linear combination of the radial basis (kernel) functions computed by the kernel units. Each kernel unit has a localized receptive field. The basis functions of the kernel layer are provided with the clustering centers which have statistical significance. Basis function can be chosen with respect to practical needs, the most widely used basis function is the simple Gaussian function in which the activation (*Oi)* of kernel unit *i* is calculated by the following formula:

$$o_j(x) = e^{\frac{-(x-m_j)\bullet(x-m_j)}{2s_j^2}}$$

where x is the input vector, μ_j is the vector determining the center of the basis function associated with kernel unit j, and s^2 is the normalization factor. The output value in the kernel unit lies between 0 and 1. The closer is an input to the center of the Gaussian function, the larger the response of the unit becomes. The normalization factor s^2 represents a measure of the spread of the data around the cluster center associated with the kernel unit. It is usually determined by the average distance between the cluster center and each training instance (point) around the center. The activation level O_j of unit j in the output layer is determined by the linear combination:

$$o_j = \sum_{i=1}^{n} w_{ji} o_i$$

where w_{ji} is the weight from kernel unit i to output unit j. In the output layer, the value of an unit is obtained through a linear combination of the nonlinear outputs from the kernel layer.

Therefore, RBF networks can be regarded as a bottom up approach to data classification by treating the design of a ANN as an approximation (curve-fitting) problem in high-dimensional space. For the problem of classification, RBF networks can determine how closely a given input is to the center of a kernel by the response of the corresponding kernel unit. If only a single kernel unit is employed, the decision region is simply circular. From this perspective, the RBF network is suitable for implementing an efficient classification model. Using a set of nonlinear basis functions, a RBF network is capable of approximating very arbitrary mapping relationship.

In addition, RBF networks can overcome the problems of slow training speed and convergence to local minimum. Leaning in RBF networks is essentially the search of a surface that provides the best fit (in some statistical sense) to the training data in multidimensional space. The learning process of the RBF network can be divided into two stages: learning in the kernel layer followed by learning in the output layer. Typically, the process of learning in the kernel layer is to determine the status of the units in the kernel layer, and it is usually performed by unsupervised clustering method. The supervised methods like Least Mean Square (LMS) algorithmare used for learning from the kernel layer to the output layer. Generally, K-means is employed as the clustering algorithm to determine the status of units in the kernel layer (Rolloet et al. 1998). Although k-means is a rapid and simple method for cluster partitioning in the feature space, it is difficult to determine the cluster centers with suitable scale. Therefore, other flexibly clustering algorithms have been demonstrated be able to overcome the aforementioned shortcomings and might perform better in determining the status of units in the kernel layer of a RBF network. In this study, the adaptive resonance theory (ART) is employed for the clustering process (Grossberg 1976), and it is essentially a cluster discovery model useful for pattern recognition and classification (Carpenter and Grossberg 1988). The ART provides a solution to the stability-plasticity dilemma during the design process of learning systems, and it has two useful properties: real-time learning and self-organization. It has been demonstrated that ART is more sensitive to data noise than other conventional clustering methods such as the K-means algorithm and ISODATA.

2.3 The Rule-based inference and evidence combination

Neural computation model is a simulation of human vision with low level of intelligence. An intelligent pattern recognition system should be able to proces s higher level of knowledge which is often symbolic in nature. A suitable integration of both will enhance the simulation of image understanding with remotely sensed data. Integration of ancillary data or knowledge in image classification has been shown to be effective in enhancing discrimination and classification accuracy (Eiumnoh and Shrestha 2000). In the present RBF classification model, geographical data or knowledge serves as ancillary information to improve the classification. Terrain features and their derived elements, such as slope and aspect, are integrated with spectral information to determine the final pattern of distribution. There are several ways in which geographical data and expert knowledge can be captured and represented by a knowledge-based analysis system . In this study, we use the simplest and perhaps the most common approach to symbolic knowledge representation, namely the production rules, to take into account the geographical knowledge of land covers. The captured rules are mostly fuzzy in meaning and uncertain in belief, i.e., consisting of two major types of uncertainty, imprecision and randomness in knowledge representation and inference. The uncertainty of rule is represented by a notion of probability taking the following format:

IF(condition), **THEN**(conclusion), **PF**(Probability Factor)

where *PF* reflects the degree of uncertainty, and *PF* ? [0, 1]. When PF is equal to 0, the rule is absolutely incredible. When *PF* is equal to 1, the conclusion of the rule is absolutely credible. When 0<PF<1, the conclusion is credible to a certain degree. To allow for imprecision in probabilistic statement, linguistic hedges can be used to modify PF.

If we treat the classification results obtained from both the RBF network and the rule-based inference as evidence leading to the final classification of land covers, then we need some methods of evidence combination to integrate the initial results to derive the final classification result. Among different methods for evidence combination, the Dempster-Shafer theory is adopted in this study. According to the Dempster-Shafer (D-S) theory of evidence combination, the final vector of probability (p_3) can be determined by the technique of orthogonal summation (denoted by ?) of p_1 (Vector of probability obtained from the RBF network) and p_2 (Vector of probability obtained from the rule-based inference) as follows:

$$p_3(G) = p_1 \oplus p_2(G) = \frac{\sum_{x \cap y} p_1(x) \bullet p_2(y)}{1 - \sum_{x \cap y = f} p_1(x) \bullet p_2(y)}$$

where *G* is a subset which represents the category, f is the empty set, and $p_3(G)$ is the partial PF attributed to *G* from the final *PF*.

3. AN APPLICATION

3.1 The study area and data used

As an evaluation, the RBF model was applied to classify land covers from TM image. In this application, experiments were conducted using TM image with 6 non-thermal bands (1-7). The study area covers the Yuenlong region, northwest of Hong Kong (Figure 2). The size of the sub image cut out from the whole image is 600 rows by 600 columns, covering about 3200km^2. According to the survey of the study area and with the vision interpretation of the corresponding data, there are 12 main types of land covers: C1—Sea; C2—Beach; C3—Inland Water; C4—Wet Land;

Figure 2. The TM Image of the study area

C5—Mangrove; C6—Urban; C7—Concrete Land; C8—Baren Land; C9—Green Land; 10—Forest; C11—Hill Grass; C12—Rock Grass

where water (C1,C3), building area (C6,C7), vegetation area (C9,C10,C11), water (C1,C3) and shadow of building (C6, C7) can not be easily separated because of their closeness in spectral characteristics. The knowledge-integrated RBF model is thus used for the task.

3.2 Results and Discussion

The 4 dimensional input vector for the RBF network is A = (PCA1, CH4, CH5, CH7), where PCA1 is the principle component of PCA transformation of visible bands (CH1-Blue, CH2-Green, CH3-Red) for dimension reduction. A total of 1700 training samples were selected through visual interpretation of the scenes by comparing with a land-use map. In the training phase, the data sets include 1700 training sample data and 800 test sample data.

Firstly, the RBF network with a kernel layer size of 120 and learning rate? of 0.01, was trained by the training sample data. The test error matrix was then obtained. The training time of the RBF network was about 50 seconds, and the overall accuracy is 90.17%. Meanwhile, the maximum likelihood classifier (MLC) and BPNN were also applied to the same data sets. The obtained structure of the BPNN is of three layers with 4 input nodes, 24 hidden nodes and 12 output nodes. The overall accuracy of the MLC is 85.25%, and that of BPNN is 89.92%. However, the learning time of the BPNN is about 1200 seconds after about 6,500,00 iterations. Comparing the three classifiers, we reach the following conclusions:

(i) Training time of the RBF network is less than that of the BPNN and the former attains higher accuracy of classification. (ii)ANN classifiers are distribution-free, and have more capability to separate the categories of mixture distribution in the feature space than conventional parametric statistical classifiers. Therefore, the RBF network yields the most effective classification both in the learning phase and the test phase.

To select the reasonable number of units in the kernel layer of the RBF network is a key to the success of the classification. In this study, different numbers of units in the kernel layer, including 30, 40, 50, 60, 90, 125, 160, 200, and 250, are respectively selected. In other words, the patterns in the feature space are partitioned into different areas by the clustering method. The results indicate that the overall accuracy can be improved by increasing the size of the kernel layer, but the computation overhead also increases. However, when size of the kernel layer increases to a certain magnitude, the overall accuracy levels off.

Better classification can be achieved if suitable geographical knowledge can be integrated into the RBF model. As an evaluation, knowledge established in terms of DEM and slope is used. For examples, it is impossible to have sea water distributed in places with DEM being higher than 0; hill grass area generally distributes on hill tops with DEM higher than 100m; the green forest land , such as urban park, should only distribute on mild plains around urban areas; and etc. The knowledge described above is represented as rules and was employed to classify land covers in the study area by the knowledge-integrated RBF model.

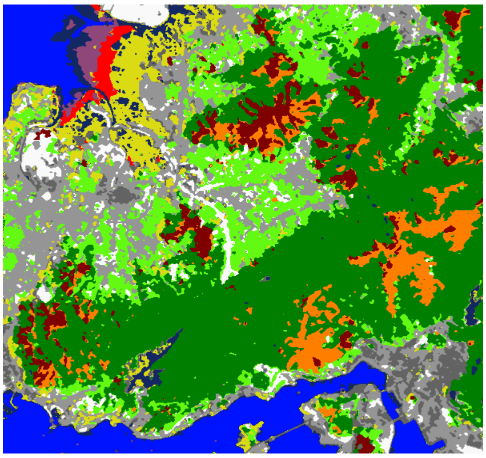

Figure 3. Land cover map obtained by the knowledge-integrated RBF model.

The same test samples were used to assess the overall accuracy of the knowledge-integrated RBF classification model. The results of the test indicate that the overall accuracy of the knowledge-integrated RBF model evidently increases to 93.17% in comparison with the 90.17% achieved by the RBF network. Furthermore, the results obtained by the knowledge-integrated RBF model are visually more nature, especially the distribution of land covers such as wetland, rock-grass and inland water. It should be noted that we had only used very few coarse domain specific knowledge in the integrated model. Better performance is expected if finer knowledge could be integrated into the classification process.

4. CONCLUSIONS

We have proposed in this study a knowledge-integrated RBF model for remote sensing classification. The integrated model takes advantage of the efficient classification by the RBF network and the geographical knowledge-based classification by the rule-based inference. It also provide a means to combine results obtained from neural networks with low level of intelligence similar to human vision and results derived from inference with geographical knowledge similar to human thought process via techniques of evidence combination. The effectiveness of such an integration has been demonstrated by the experiment. It is evident in the experiment that classification accuracy can be enhanced even though very coarse geographical knowledge derived from DEM and its derivatives is integrated with the RBF network. To further improve knowledge-integrated neural network model in general and the integrated RBF model in particular, it is essential to study the structures of various neural network models and the schemes of various knowledge representations so that a suitable integration can be derived.

References

[1] Atkinson, P. M., and Tatnall, A. R. L., 1997, Neural networks in remote sensing, International Journal of Remote Sensing, 18(4): 699-709

[2] Bishop, C. M., 1995, Radial basis functions. In "Neural Networks for Pattern Recognition", Clarendon Press, Oxford, New York.

[3] Bruzzone, L., and Prieto, D. F., 1999, A technique for the selection of kernel-function parameters in RBF neural networks for classification of remote-sensing images, IEEE Transactions on Geoscience and Remote Sensing, 37(2): 551-559

[4] Eiumnoh, A., and Shrestha, R. P., 2000, Application of DEM Data to Landsat Image Classification: Evaluation in a Tropical Wet-Dry Landscape of Thailand, Photogrammetric Engineering & Remote Sensing, 66(3): 297-304

[5] Foody, G. M., 1995, Using prior knowledge in artificial neural network classification with a minimal training set. International Journal of Remote Sensing, 16, 301-312.

[6] Gong, P., 1996, Integrated Analysis of Spatial Data from Multiple Sources: Using Evidential Reasoning and Artificial Neural Network Techniques for Geological Mapping, Photogrammetric Engineering & Remote sensing, 62[5]: 513-523

[7] Grossberg, S., 1976, Adaptive pattern classification and universal recording I: parallel development and coding neural feature detectors. Biological Cybernetics, 23, 121-134.

[8] Medsker, L. R., 1994, Hybrid neural network and expert systems, Kluwer Academic Publishers

[9] Murai, H. and Omatu, S., 1997, Remote Sensing image analysis using a neural network and knowledge-based processing, International Journal of Remote sensing, 18(4): 811-828

[10] Peddle, D. R., 1995, Knowledge Formulation for Supervised Evidential Classification, Photogrammetric Engineering & Remote Sensing, 61[4]: 409-417

[11] Powell, M. J. D., 1987, Radial Basis Functions for multivariable interpolation: r review. In Algorithms for Approximation of Functions and Data, edited by J. C. Mason, and M. G. Cox, pp. 143-167, Oxford University Press, Oxford

[12] Rollet, R., Benie, G.B., Li, W., and Wang, S., 1998, Image Classification algorithm based on the RBF neural network and K-means, International Journal of Remote Sensing, 19(15):3003-3009