APPLICATION OF MLP AND RBF NETWORKS TO CLOUD DETECTION

W.D. Zhang, M.X. He and M.W. Mak[†]

Ocean Remote Sensing Laboratory of Ministry of Education of China, Ocean Remote Sensing Institute (ORSI), Ocean University of Qingdao, China. wdzhang_cn@21cn.com, mxhe@ns.qd.sd.cn
[†]Center for Multimedia Signal Processing, Dept. of Electronic and Information Engineering, The Hong Kong Polytechnic University, Hong Kong, China. enmwmak@polyu.edu.hk

ABSTRACT

This paper compares the performances of multi-layer perceptrons (MLPs) and radial basis function (RBF) networks on detecting clouds in NOAA/AVHRR images. The main results show that the RBF networks are able to handle complex atmospheric and oceanographic phenomena while the conventional rule-based systems and MLPs can not. In particular, the experimental evaluations show that the RBF networks can converge to global minima while the MLPs can only achieve this occasionally, and that classification errors made by the RBF networks decrease dramatically when the number of basis functions increases. In addition, these errors are almost identical when the number of basis functions reaches a threshold. Only on a few rare occasions when the backpropagation algorithm attains an optimal solution and the classification errors made by the MLPs be comparable to (but still larger than) the ones made by the RBF networks. However, the results show that achieving such optimal solutions is difficult. It is, therefore, concluded that the RBF networks are better than the MLPs for cloud detection.

1. INTRODUCTION

Highly accurate and automatic detection of clouds in satellite multi-spectral data is fundamental to climatological and oceanographic applications [1]. In these applications, cloud detection can be considered as an image segmentation problem in which images are classified into two classes: clear and cloudy.[1]

Conventional rule-based cloud detection systems [2] have been extensively used in the past. Some of these systems have been used in operational applications [3] while others are still under development [4]. The accuracy of rule-based systems, however, can be affected by various atmospheric and oceanic conditions, which might lead to misleading detection results in some situations such as coastal and high aerosol concentration areas. Neural classifiers, because of their adaptive learning capabilities, offer an attractive alternative for cloud detection. Most of the previously proposed methodologies are based on the Multi-layer Perceptrons (MLPs) trained by the error back propagation (EBP) algorithm using spectral and textural characteristics as features [5, 6]. However, MLPs have their own limitations. For example, the slow convergence of the EBP learning algorithm and its potential of converging to local minima may render MLPs impractical in real applications. Radial basis function (RBF) networks overcome these problems by means of a fast training algorithm such that locally optimal solutions are avoided. Such characteristics and their intrinsic simplicity are expected to make RBF networks a better candidate for cloud detection.

This paper is organized as follows. Section 2 provides a brief introduction to neural classifiers (including MLPs and RBFs) and explains how they can be applied to cloud detection. Experimental evaluations are set forth in Section 3. The performances of the RBF networks and the MLPs in cloud detection are compared in Section 4. Finally, concluding remarks are drawn in Section 5.

2. NEURAL CLASSIFIERS

2.1. Multi-layer Perceptrons

Multi-layer Perceptrons are feed-forward neural networks with one or more hidden layers [7]. Knowledge is encoded in the synaptic weights interconnecting the processing nodes. The most popular learning algorithm for MLPs is the error backpropagation (EBP) algorithm [7].

MLPs have been widely studied and applied to solve cloud detection problems (see e.g. [5]). It is well known that MLPs are able to construct arbitrary decision boundaries [8], and therefore are applicable to cloud detection. In the context of cloud detection, the number of inputs to the networks is determined by the number of features in a pixel. Similarly, the number of outputs is equal to the number of classes. The number of hidden nodes is a free parameter and its value depends on the complexity of the classification problem.

2.2. RBF Neural Networks

RBF networks are three-layer feed-forward networks with a single hidden layer [9]. They have been applied to a wide range of pattern recognition problems [10]. When used as pattern classifiers, RBF networks represent the posterior probabilities of the training data by a weighted sum of Gaussian basis functions with diagonal covariance matrices. The kth output ($k = 1, \ldots, K$) of an RBF network with I inputs and J function centers has the form

$$y_k(\vec{x}_p) = w_{k0} + \sum_{j=1}^{J} w_{kj}\phi_j(\vec{x}_p) \quad p = 1, \ldots, N \qquad (1)$$

where $\phi_j(\vec{x}_p) = \exp\left\{ -\sum_{i=1}^{I} \frac{(x_{pi} - \mu_{ji})^2}{2\sigma_j^2} \right\}$, $\vec{x}_p$ is the pth input vector, $\vec{\mu}_j$ and σ_j are the mean vector and width of

[1]Hereafter, we use the terms 'clear' and 'cloudy' to refer to the ocean and cloud regions of a satellite image, respectively.

the jth basis function respectively, w_{k0} is a bias term and w_{kj} is the output weight connecting the jth basis function with the kth output. Features' information is stored in the centers $\{\vec{\mu}_j\}$ and the widths $\{\sigma_j\}$ of the basis functions, and the weights $\{w_{kj}\}$ represent the relative importance of the basis functions in response to an external input.

The mean vectors and the widths of an RBF network can be estimated by the K-means algorithm and the K-nearest neighbors algorithm, respectively [9]. The output weights can be determined by a least squares approach using the technique of singular value decomposition.

To model the data variability caused by atmospheric and oceanic conditions, the network topology is divided into two parts. More precisely, two sets of basis functions are used to model the clear and cloudy data, and two output nodes are used for discrimination. For the cloudy class, the first output node is trained to output a '1' and the second to output a '0', and vice versa for the clear class. Similar to MLPs, the numbers of nodes in the input and output layers are defined by the number of feature components and the number of classes, respectively. Therefore, we only need to determine the number of nodes in the hidden layer. However, no proven algorithms are available to determine the optimal number of nodes in the hidden layer.

2.3. Threshold Selection Techniques

During recognition, the feature vectors derived from the feature generation procedure (see Section 3.1) form a vector sequence $X = \{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_T\}$. For each $\vec{x}$ in X, the network produces two outputs: $y_k(\vec{x})$, $k = 1, 2$. The outputs are then normalized by a softmax function as follows:

$$z_k = \frac{e^{\tilde{y}_k(\vec{x})}}{\sum_{r=1}^{2} e^{\tilde{y}_r(\vec{x})}} \quad k = 1, 2 \tag{2}$$

where $\tilde{y}_k(\vec{x}) = \frac{y_k(\vec{x})}{P(C_k)}$ represents the scaled outputs and $P(C_k)$ is the prior probability of class C_k. For MLPs, no output normalization is necessary because the outputs are always bounded between 0.0 and 1.0 (i.e. they are normalized internally). Therefore, for MLPs, we used the scaled output

$$z_k = \frac{y_k(\vec{x})}{p(C_k)} \quad k = 1, 2 \tag{3}$$

for classification.

We also need to map the continuous network outputs onto discrete class labels. For networks with two outputs, one for the cloudy class and one for the clear class, a typical class labeling rule is

$$l(\vec{x}) = arg \max_{k=1}^{2} \{z_k(\vec{x})\} \tag{4}$$

where $z_k(\vec{x})$ are either the normalized outputs of the RBF networks (2) or the scaled outputs of the MLPs (3). Note that this rule makes use of the notion that the network's outputs are the estimate of the *a posteriori* probabilities, i.e., $y_k(\vec{x}) \simeq p(C_k|\vec{x})$, where $P(\cdot)$ denotes *a priori* probability and C_k denotes the kth class [8]. Here, the decision criterion can be written as:

$$\text{If } z_1 - z_2 \begin{cases} > \zeta & \text{then } \vec{x} \in \text{cloudy class} \\ \leq \zeta & \text{then } \vec{x} \in \text{clear class} \end{cases} \tag{5}$$

where $\zeta \in [-1, 1]$ is a decision threshold. A decision is made for each input vector, and the error rate is the proportion of incorrect labeling decisions to the total number of decisions.

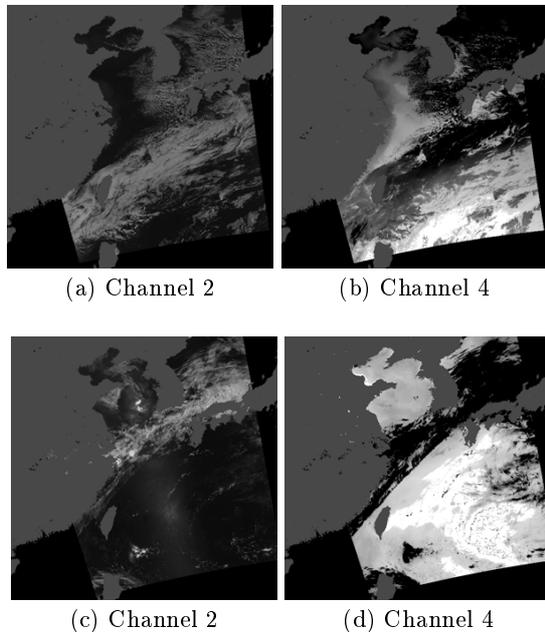(a) Channel 2   (b) Channel 4

(c) Channel 2   (d) Channel 4

Figure 1: AVHRR images obtained from Channel 2 [(a) and (c)] and Channel 4 [(b) and (d)]. (a) and (b) were acquired on Jan 5, 1998 while (c) and (d) were acquired on June 28, 1998.

The optimal decision rule tells us that ζ should theoretically be zero. However, we found in a previous study [11] that the threshold ζ is nonzero if one of the classes has a larger variance than the other. In the case of cloud detection, the variance of clouds is much larger than that of the clear ocean. Therefore, the equal error threshold ζ^* that causes an equal chance of misclassifying both classes is slightly larger than zero. Therefore, the optimal value of ζ should be determined experimentally. This can be achieved by reserving some training samples (the trained models have never seen these data before) and feeding them to the trained models. The threshold ζ is adjusted until the error of misclassifying clouds is equal to that of misclassifying the clear ocean.

3. EXPERIMENTS

3.1. Data Set and Feature Generation

We evaluated an MLP-based and an RBF-based cloud detector based on data obtained from the Advanced Very High Resolution Radiometer (AVHRR) on board the NOAA-14 satellite. The AVHRR has five channels: Channel 1 to Channel 5. These channels are designed to detect signals with different wavelengths [4]. For example, clouds can be identified by their high albedo in visible channels and low brightness temperature in infrared channels. Images shown in Figs. 1 (a) and (b) were acquired on January 5, 1998 at 5:43 UTC and Figs. 1 (c) and (d) were acquired on June 28, 1998 at 5:35 UTC. These scenes are of particular interest because 1) high albedo and cold coastal water are present in images (a) and (b); 2) high aerosol concentration is present in the middle of the images (c) and (d). Conventional rule-based systems often misclassify the clear ocean as clouds in these situations (e.g. see Fig. 5 (b), where only small regions of the ocean are identified) if we do not adjust the
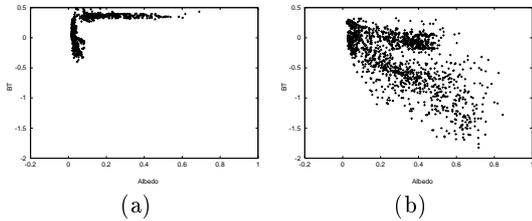
Figure 2: Scatter plots of feature vectors derived from (a) the clear ocean and (b) clouds.
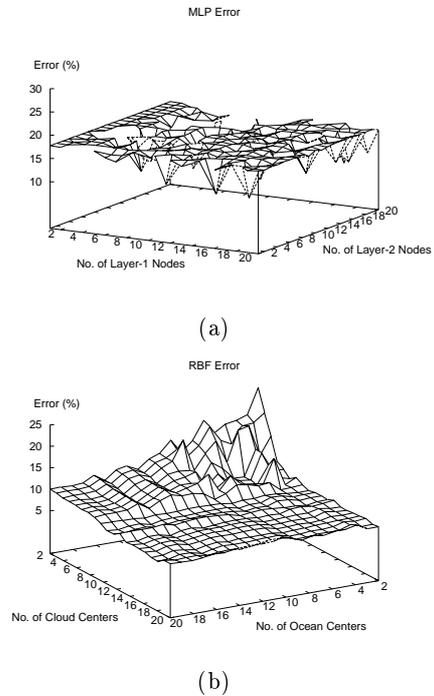

(a)

(b)

Figure 3: Classification errors (based on training data) made by (a) MLP and (b) RBF networks with various number of nodes in their hidden layers.

decision thresholds to accommodate the variations.

Because no ideal cloud detection methods are presently available, testing data are often obtained from manual segmentation performed by human operators. Similarly, our training and testing data were also segmented by human experts. The main goal of this study was to compare the performance of the RBF networks with that of the MLPs in cloud detection. Therefore, only two spectral channels, namely Channel 2 (visible) and Channel 4 (IR), were used. These channels were selected because 1) they carried most of the clouds' characteristics, e.g. albedo in Channel 2 and temperature signals in Channel 4; and 2) the entire decision space could be displayed on a 2-D plane, which helps us investigate the internal representations of these two classifiers. Fig. 2 is a scatter plot of the feature vectors. The horizontal-axis is the albedo computed from Channel 2 and the vertical-axis is the brightness temperature (BT) computed from Channel 4. For ease of comparison, the features have been scaled according to [5]. We can see from Fig. 2 (a) that the ocean data generally lie in the upper-left corner of the feature space. They generally exhibit low albedo and high BT. However, there are exceptions: 1) some clear ocean data exhibit low BT but high albedo, which are mainly caused by winter images and 2) some clear data have high BT and high albedo which are caused by summer images with high aerosol concentration. Clouds, on the other hand, typically have low BT and high albedo, but they exhibit high variability and therefore are scattered over a wide area of the feature space.

3.2. Network Training

The data were divided into three sets: training set, validation set, and test set. One-twentieth of the total number of pixels (approximately 170,000 pixels derived from the clear ocean and 170,000 pixels from the clouds) were used as the training set. The validation set, which was the same size as the training set, was formed by drawing pixels randomly from the remaining data. The validation set was used for terminating the EBP training and for decision threshold determination. The rest of the data constitutes the test set.

In this study, we investigated MLPs with one and two hidden layers where the numbers of nodes in the hidden layer(s) varied from 1 to 20 in steps of one node. For RBF networks, the number of nodes representing the clear class (denoted as ocean centers) and the cloudy class (denoted as cloud centers) varied from 2 to 20 in steps of one node.

4. RESULTS AND DISCUSSION

Fig. 3 depicts the training classification errors made by networks with various numbers of hidden nodes, cloud cen-

ters and ocean centers. The classification errors were determined by presenting the training patterns to the trained models and counting the number of incorrect decisions made by the models. Due to the large amount of training patterns, the classification errors obtained by presenting the training set, validation set or test set to the models are almost identical. Therefore, the classification errors based on the validation set and test set are not shown. We can see from Fig. 3 that for MLPs, only a few network configurations achieve an error rate lower than 10%, meaning that it is difficult to find the best network configuration. For those networks with a relatively small error rate, it may well be the case that their initial synaptic weights are already very close to their global optimum. However, it is well known that finding such initial weights is very difficult in practice.

For the RBF networks, we can observe from Fig. 3 that the error rate is very high when the number of nodes is less than 4. The errors, however, decrease dramatically when the number of nodes in each class is larger than 4. The errors made by the RBF networks are generally less. For example, the lowest test error rate that we obtained was 6.4% for the RBF networks and 7.2% for the MLPs.

Unlike the MLPs, RBF networks with numbers of nodes larger than 20 (10 for each class) have almost identical errors. These results suggest that the optimal number of nodes for the task concerned is about 20. We also found that the RBF networks generally require more cloud centers to model the cloudy class than the clear class because of the large disparity of the cloud data in the feature space.

Significant differences between MLPs and RBF networks can also be observed from their decision boundaries. As shown in Fig. 4, the MLP uses a straight line to separate the two classes (see Fig. 2 for the real boundary between
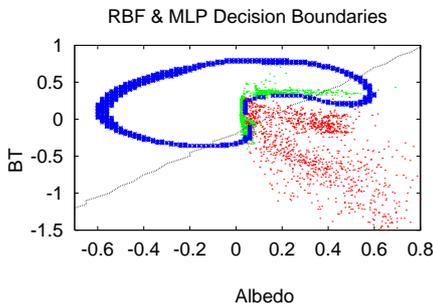
RBF & MLP Decision Boundaries

Figure 4: Decision boundaries formed by an RBF network (black curve) and an MLP (grey line). See Fig. 2 for the real boundary between these two classes.
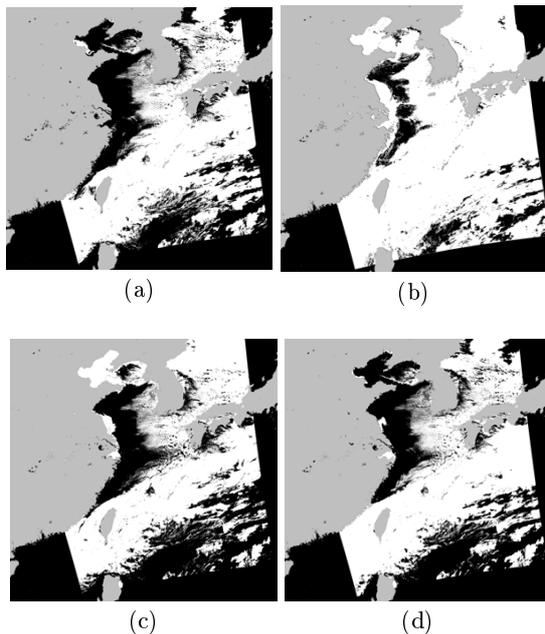

(a)   (b)

(c)   (d)

Figure 5: Clouds detected by (a) human, (b) a rule-based system, (c) an MLP, and (d) an RBF network. Clouds are represented in white, land in gray and clear ocean in black.

the two classes). The RBF networks, on the other hand, produces a curve to separate the two classes almost perfectly.

Fig. 5 shows the cloud detection results of the satellite imagery shown in Fig. 1 (a). A hand labeled image (Fig. 5 (a)) is provided as a reference for comparison. The result of the rule-based system (Fig. 5 (b)) was obtained by a commercial software package without any threshold adjustments. We can see that the rule-based system falsely detects the clouds in the Bohai sea, along the coastline and along the Kuroshio edge (bottom-right portion). Fig. 5 (c) shows that the MLP is better than the rule-based system, but it also fails to detect the clouds in the Bohai sea and the upper-right region where very cold water exists. Fig. 5 (d) demonstrates that the RBF network achieves the most promising results and the segmented image is closest to the one obtained by human experts (Fig. 5 (a)).

5. CONCLUSION

This paper presents a scheme to model clear and cloudy data using RBF networks. The performances of the RBF networks and the MLPs in cloud detection has been compared. The experimental evaluations show that the locally-tuned nature of RBF networks is very useful for detecting clouds. We have shown that the RBF networks generally achieve a lower error rate than the MLPs. Although the MLPs can achieve a low detection error rate if an optimal architecture is obtained, such architecture is not easy to find in practice.

6. REFERENCES

[1] R. A. Schiffer and W. B. Rossow. The international satellite cloud climatology program (ISCCP): the first project of the world climate research program. *Bulletin of American Meteorological Society*, 64:779–784, 1983.

[2] R. W. Saunders and K. T. Kriebel. An improved method for detecting clear sky and cloudy radiances from AVHRR data. *Int. Journal of Remote Sensing*, 9:123–150, 1988.

[3] E. P. McClain W. G. Pichel and C. C. Walton. Comparative performance of AVHRR-based multichannel sea surface temperatures. *Journal of Geophisical Research*, 90(C6):11587–11601, 1985.

[4] L. L. Stowe, P. A. Davis, and E. P. McClain. Scientific basis and initial evaluation of the CLAVR-1 global clear/cloud classification algorithm for advanced very high resolution radiometer. *Journal of atmospheric and oceanic technology*, 16:656–681, 1999.

[5] S. Yhann and J. J. Simpson. Application of neural networks to AVHRR cloud segmentation. *IEEE Trans. on Geo-science and Remote Sensing*, 33(3):590–604, 1995.

[6] G. S. Pankiewicz. Pattern recognition techniques for the identification of cloud and cloud systems. *Meteorological Application*, 2:257–271, 1995.

[7] D. E. Rummelhart, J. L. McClell, and et al. *Parallel Distributed Processing, Explorations in the Microstructure of Cognition, Vol. 1: Foundations*. MIT Press, Cambridge, MA, 1987.

[8] R. P. Lippmann. An introduction to computing with neural nets. *IEEE ASSP Magazine*, 4:4–22, 1987.

[9] J. Moody and C. J. Darken. Fast learning in networks of locally tuned processing units. *Neural Computation*, 1:281–294, 1989.

[10] M. W. Mak and S. Y. Kung. Estimation of elliptical basis function parameters by the EM algorithm with application to speaker verification. *IEEE Transactions on Neural Networks*, 11(4):961–969, 2000.

[11] W. D. Zhang, K. K. Yiu, M. W. Mak, C. K. Li, and M. X. He. A prior threshold determination for phrase-prompted speaker verification. In *Proc. of Eurospeech'99*, volume 2, pages 1023–1026, Budapest, Hungary, Sept. 1999.